PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/54, 9/12, C07K 16/40, C12Q 1/68, A61K 38/45, A01K 67/027, G01N 33/50, 33/573

A2

(11) International Publication Number: WO 99/07854

(43) International Publication Date: 18 February 1999 (18.02.99)

(21) International Application Number: PCT/US98/16640

(22) International Filing Date: 11 August 1998 (11.08.98)

(30) Priority Data:
08/909,354 11 August 1997 (11.08.97) US

(71) Applicant *(for all designated States except US)*: ONTOGENY, INC. [US/US]; 45 Moulton Street, Cambridge, MA 02138 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PANG, Kevin [US/US]; 53 Hill Road #10, Belmont, MA 02178 (US). MIAO, Ningning [US/US]; 671 Washington Street, Brookline, MA 02146 (US). BARKER, Douglas, D. [US/US]; 47 Louise Road, Belmont, MA (US).

(74) Agents: VINCENT, Matthew, P. et al.; Foley, Hoag & Eliot, LLP, One Post Office Square, Boston, MA 02109 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: SERINE/THREONINE KINASE, AND USES RELATED THERETO

MOTIF STUDY IN KINASE DOMAIN THAT PREDICT KINASE SPECIFICITY

KINASE	SUBDOMAIN	FUNCTION
1.GXGXXG	I	ATP BINDING SITE
2.HRDLKSKN	VIB	IN MOST OF STK
3.HRDLAARN	VIB	IN MOST OF TK
4.GTKRYMAPE	VIII	IN MOST OF STK
5.XP(IV)(K/R)W(T/M)	VIII	IN MOST OF TK
6.DFG	VII	IN ALL OF PK

GXGXXG: ATP BINDING SITE NOT MATCH ON 9-2-1.9kb

TABLE 2 COMPARE WITH OTHER STKR MEMBERS

KINASE	SUBDOMAIN VIB	VIII
1.9-2-1.9kb	HRDLKPEN	GTPCWMAPE
2.ActR-II	HRDLKSKN	GTRRYMAPE
3.ActR-IIB	HRDFKSKN	GTRRYMAPE
4.TBR-II	HRDLKSSN	GTARYMAPE
5.ALK-I	HRDFKSRN	GTKRYMAPE

(57) Abstract

We describe here a new class of serine/threonine kinase receptors, called "*pan-s/tk*". The sequence of exemplary *pan-s/tk* genes indicates that they encode receptor type serine/threonine kinases with a single kinase domain.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Serine/Threonine Kinase, and Uses Related Thereto

Background of the Invention

Receptors on the surfaces of cells transmit information into the cytoplasm to effect appropriate responses to extracellular signals. Protein phosphorylation has been extensively characterized as a major mechanism of transducing signals within cells. Many signal transduction pathways involved in the control of cell proliferation and differentiation originate with transmembrane receptors containing cytoplasmic protein kinase domains. For instance, there are many kinds of cytokine and growth factor receptors that belong to different receptor families. Most of these receptors function as: (1) receptor tyrosine kinase and tyrosine-kinase-associated receptor; (2) receptor serine/threonine kinase; (3) G-protein linked receptor. Following binding with an extracellular factor and activation, the receptors trigger different cascade of intracellular protein phosphorylation to transduction signals, thereby altering the cell's pattern of gene expression and leading to biological effects.

Although much of the research done has focused on receptor tyrosine kinases (see, for example, Fanti et al. (1993) *Annu. Rev. Biochem.* 62:453), receptor serine-threonine kinases (RSTKs) have been identified as well. Receptor serine-threonine kinases mediate inhibitory as well as stimulatory signals for growth and differentiation by binding to a variety of different extracellular factors. For instance, certain RSTKs bind to members of the transforming growth factor beta (TGF-β) superfamily (Massagub et al. (1992) *Cell* 69:1067; Attisano et al. (1992) *Cell* 68:97; and Lin et al. (1993) *Trends Cell. Biol* 3:14) and all known receptor-like kinases from higher plants (Walker (1993) *Plant* 3:451 (1993); Chang et al. (1992) *Plant Cell* 4:1263; Stein et al. (1991) *PNAS* 88:8816; Tobias et al. (1992) *Plant Physiol.* 99:284). RSTKs which have been isolated so far display wide expression patterns in peripheral tissues and in the nervous system.

Summary of the Invention

The present invention relates to the discovery of a new class of the receptor serine/threonine kinase receptors (RSTKs), referred to herein as *pan-s/tk* (for pancreatic) receptors.

In general, the invention features isolated *pan-s/tk* polypeptides, preferably substantially pure preparations of the subject *pan-s/tk* polypeptides. The invention also provides recombinantly produced *pan-s/tk* polypeptides. In preferred embodiments the polypeptide has a biological activity including the ability to phosphorylate a serine or threonine residue of an intracellular protein or peptide substrate. However, *pan-s/tk*

polypeptides which specifically antagonize such activities, such as may be provided by truncation mutants or other dominant negative mutants, are also specifically contemplated.

The *pan-s/tk* proteins of the present invention can be characterized as including one or more of the following domains/motifs: an extracellular domain, ee.g., which mediates ligand binding, a transmembrane domain, and an intracellular domain including a kinase domain. The protein may also include a secretion signal sequence, and (optionally) glycosylated amino acid residues.

In one embodiment, the polypeptide is identical with or homologous to a *pan-s/tk* protein represented in SEQ ID No. 2, 4, 6 or 9. Related members of the *pan-s/tk* family are also contemplated, for instance, a *pan-s/tk* polypeptide preferably has an amino acid sequence at least 65%, 70%, 75% or 80% homologous to the polypeptide represented by SEQ ID No. 2, 4, 6 or 9, though polypeptides with higher sequence homologies of, for example, 85, 90% and 95% or are also contemplated. In a preferred embodiment, the *pan-s/tk* polypeptide is encoded by a nucleic acid which hybridizes under stringent conditions with a nucleic acid sequence represented in SEQ ID No. 1, 3, 5, 7 or 8. Homologs of the subject *pan-s/tk* proteins also include versions of the protein which are resistant to post-translation modification, as for example, due to mutations which alter modification sites (such as tyrosine, threonine, serine or aspargine residues), or which prevent glycosylation of the protein, or which prevent interaction of the protein with extracellular ligands or with intracellular proteins involved in signal transduction.

The *pan-s/tk* polypeptide can comprise a full length protein, such as represented in SEQ ID No. 2, 4, 6 or 9, or it can comprise a fragment corresponding to one or more particular motifs/domains, or to arbitrary sizes, e.g., at least 5, 10, 25, 50, 100, 150 or 200 (preferably contiguous) amino acids in length. In preferred embodiments, the *pan-s/tk* polypeptide includes a sufficient portion of the extracellular domain to be able to specifically bind to a *pan-s/tk* ligand. Truncated forms of the protein include, but are not limited to, soluble extracellular domain fragments, soluble intracellular domains including the kinase domain, and membrane-bound forms of either which include the transmembrane domain.

The subject proteins can also be provided as chimeric molecules, such as in the form of fusion proteins. For instance, the *pan-s/tk* protein can be provided as a recombinant fusion protein which includes a second polypeptide portion, e.g., a second polypeptide having an amino acid sequence unrelated (heterologous) to the *pan-s/tk* polypeptide, e.g. the second polypeptide portion is glutathione-S-transferase, e.g. the second polypeptide portion is an enzymatic activity such as alkaline phosphatase, e.g. the second polypeptide portion is an epitope tag.

In yet another embodiment, the invention features a nucleic acid encoding a *pan-s/tk* polypeptide, which has the ability to modulate, e.g., either mimic or antagonize, at least a portion of the activity of a wild-type *pan-s/tk* polypeptide. An exemplary *pan-s/tk*-encoding nucleic acid sequence is represented by SEQ ID No. 1, 3, 5, 7 or 8.

In another embodiment, the nucleic acid of the present invention includes a coding sequence which hybridizes under stringent conditions with the coding sequence designated in SEQ ID No. 1, 3, 5, 7 or 8. The coding sequence of the nucleic acid can comprise a sequence which is identical to a coding sequence represented in of SEQ ID No. 1, 3, 5, 7 or 8, or it can merely be homologous to that sequences. In preferred embodiments, the nucleic acid encodes a polypeptide which specifically modulates, by acting as either an agonist or antagonist, one or more of the bioactivities of a wild-type *pan-s/tk* polypeptides.

Furthermore, in certain preferred embodiments, the subject *pan-s/tk* nucleic acid will include a transcriptional regulatory sequence, e.g. at least one of a transcriptional promoter or transcriptional enhancer sequence, which regulatory sequence is operably linked to the *pan-s/tk* gene sequence. Such regulatory sequences can be used in to render the *pan-s/tk* gene sequence suitable for use as an expression vector. This invention also contemplates the cells transfected with said expression vector whether prokaryotic or eukaryotic and a method for producing *pan-s/tk* proteins by employing said expression vectors.

In yet another embodiment, the nucleic acid hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides of either sense or antisense sequence of SEQ ID No. 1, 3, 5, 7 or 8; though preferably to at least 25 consecutive nucleotides; and more preferably to at least 40, 50 or 75 consecutive nucleotides of either sense or antisense sequence of SEQ ID No. 1, 3, 5, 7 or 8.

Yet another aspect of the present invention concerns an immunogen comprising a *pan-s/tk* polypeptide in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for a *pan-s/tk* polypeptide; e.g. a humoral response, e.g. an antibody response; e.g. a cellular response. In preferred embodiments, the immunogen comprising an antigenic determinant, e.g. a unique determinant, from the protein represented by SEQ ID No. 2, 4, 6 or 9.

A still further aspect of the present invention features antibodies (monoclonal, polyclonal or recombinant) and antibody preparations specifically reactive with an epitope of the *pan-s/tk* immunogen.

The invention also features transgenic non-human animals, e.g. mice, rats, rabbits, chickens, frogs or pigs, having a transgene, e.g., animals which include (and preferably express) a heterologous form of a *pan-s/tk* gene described herein, or which misexpress an endogenous *pan-s/tk* gene, e.g., an animal in which expression of one or more of the subject *pan-s/tk* proteins is disrupted. Such a transgenic animal can serve as an animal model for

studying cellular and tissue disorders comprising mutated or mis-expressed *pan-s/tk* alleles or for use in drug screening.

The invention also provides a probe/primer comprising a substantially purified oligonucleotide, wherein the oligonucleotide comprises a region of nucleotide sequence which hybridizes under stringent conditions to at least 12 consecutive nucleotides of sense or antisense sequence of SEQ ID No. 1, 3, 5, 7 or 8, or naturally occurring mutants thereof. In preferred embodiments, the probe/primer further includes a label group attached thereto and able to be detected. The label group can be selected, e.g., from a group consisting of radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors. Probes of the invention can be used as a part of a diagnostic test kit for identifying dysfunctions associated with mis-expression of a *pan-s/tk* protein, such as for detecting in a sample of cells isolated from a patient, a level of a nucleic acid encoding a *pan-s/tk* protein; e.g. measuring a *pan-s/tk* mRNA level in a cell, or determining whether a genomic *pan-s/tk* gene has been mutated or deleted. These so-called "probes/primers" of the invention can also be used as a part of "antisense" therapy which refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridize (e.g. bind) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding one or more of the subject *pan-s/tk* proteins so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. Preferably, the oligonucleotide is at least 12 nucleotides in length, though primers of 25, 40, 50, or 75 nucleotides in length are also contemplated.

In yet another aspect, the invention provides an assay for screening test compounds for inhibitors, or alternatively, potentiators, of an interaction between a *pan-s/tk* protein and, for example, a virus, an extracellular ligand of the *pan-s/tk* protein, or an intracellular protein which binds to the *pan-s/tk* protein, e.g., a substrate of the pan-s/tk kinase activity. An exemplary method includes the steps of (i) combining a *pan-s/tk* polypeptide or bioactive fragments thereof, a *pan-s/tk* target molecule (such as a pan-s/tk ligand or a pan-s/tk substrate), and a test compound, e.g., under conditions wherein, but for the test compound, the *pan-s/tk* protein and target molecule are able to interact; and (ii) detecting the formation of a complex which includes the *pan-s/tk* protein and the target polypeptide either by directly quantitating the complex, by measuring inductive effects of the *pan-s/tk* protein, or, in the instance of a substrate, measuring the conversion to product. A statistically significant change, such as a decrease, in the interaction of the pan-s/tk and target molecule in the presence of a test compound (relative to what is detected in the absence of the test compound) is indicative of a modulation, e.g., inhibition or potentiation, of the interaction between the *pan-s/tk* protein and the target molecule.

Yet another aspect of the present invention concerns a method for modulating one or more of growth, differentiation, or survival of a cell by modulating *pan-s/tk* bioactivity, e.g., by potentiating or disrupting certain protein-protein interactions. In general, whether

carried out *in vivo*, *in vitro*, or *in situ*, the method comprises treating the cell with an effective amount of a *pan-s/tk* therapeutic so as to alter, relative to the cell in the absence of treatment, at least one of (i) rate of growth, (ii) differentiation, or (iii) survival of the cell. Accordingly, the method can be carried out with *pan-s/tk* therapeutics such as peptide and peptidomimetics or other molecules identified in the above-referenced drug screens which agonize or antagonize the effects of signaling from a *pan-s/tk* protein or ligand binding of a *pan-s/tk* protein. Other *pan-s/tk* therapeutics include antisense constructs for inhibiting expression of *pan-s/tk* proteins, and dominant negative mutants of *pan-s/tk* proteins which competitively inhibit ligand interactions upstream and signal transduction downstream of the wild-type *pan-s/tk* protein.

Another aspect of the present invention provides a method of determining if a subject, e.g. an animal patient, is at risk for a disorder characterized by unwanted cell proliferation or aberrant control of differentiation or apoptosis. The method includes detecting, in a tissue of the subject, the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a *pan-s/tk* protein, e.g. represented in SEQ ID No. 1, 3, 5, 7 or 8 or a homolog thereof; or (ii) the mis-expression of a *pan-s/tk* gene. In preferred embodiments, detecting the genetic lesion includes ascertaining the existence of at least one of: a deletion of one or more nucleotides from a *pan-s/tk* gene; an addition of one or more nucleotides to the gene, a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene; an alteration in the level of a messenger RNA transcript of the gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; a non-wild type level of the protein; and/or an aberrant level of soluble *pan-s/tk* protein.

For example, detecting the genetic lesion can include (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of a *pan-s/tk* gene or naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with the *pan-s/tk* gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and (iii) detecting, by hybridization of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion; e.g. wherein detecting the lesion comprises utilizing the probe/primer to determine the nucleotide sequence of the *pan-s/tk* gene and, optionally, of the flanking nucleic acid sequences. For instance, the probe/primer can be employed in a polymerase chain reaction (PCR) or in a ligation chain reaction (LCR). In alternate embodiments, the level of a *pan-s/tk* protein is detected in an immunoassay using an antibody which is specifically immunoreactive with the *pan-s/tk* protein.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art.

Such techniques are explained fully in the literature. See, for example, Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Methods In Enzymology, Vols. 154 and 155 (Wu et al. eds.), Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

Figure 1 is a Southern blot (.1 x SSC, 60C) with a probe derived from a *pan-s/tk* transcript. The blot indicates at least two related transcripts, corresponding to the 2.2 and 4.0 kb bands. These transcripts are detectable in samples from a range of eukaryotic organims.

Figure 2 is a Northern blot with a probe derived from a *pan-s/tk* transcript. This blot also reveals the presence of at least two *pan-s/tk* transcripts, which are present in testis, kidney, heart and brain of adult rat. We observed similar transcripts in ther Northern blots (not shown) of adult and fetal pancreas. The Northern blot shows that the kidney and brain have predominantly the 4.0kb species and very little of the 2.2kb species. The testis, pancreas and heart apparently express both the 2.2 and 4.0kb species.

Figures 3A and 3B are pictures of a whole-mount *in situ* hybridization experiment with probe derived from a *pan-s/tk* transcript. The *in situ* hybridization results show expression of *pan-s/tk* transcript(s) in developing heart, brain (forebrain), a ganglia anterior of the otic vesicle, the dorsal neural tube, the pancreatic epithelium, and early gut endoderm. Expression in the gut endoderm appears to happen early, indeed the entire doudenal epithelial appears positive, and is lost as the epithelial cells mature later in development. The expression of *pan-s/tk* may therefore represent an early marker of endoderm epithelialization. Expression is also seen in the posterior-distal aspect of the

limb, encompassing the ZPA, and moving distally towards the progress zone of the limb bud. Expression in the liver was not observed. There appears to be some early kidney expression (e14 or later) and some expression in the area of the adrenal medulla. Expression in the pancreatic epithelium increases with time in development as demonstrated by the increasing intensity of expression in the pancreas. Moreover, the *in situ* hybridization of e14 rat embryo sections (see Figure 3B) shows definite expression in a subset of pancreatic epithelial ducts. The sagittal section of the e14 embryo shows expression in the heart (possibly cardiomyocytes), kidney epithelium, dorsal root ganglia, inferior ganglion (X) of vagus nerve, choroid plexus, chochlear vestibular complex, and facial (VII) ganglion. There is also a low level expression of the gene in the mesenchyme surrounding the primary, but not secondary branches of the e14 lung. There is also *pan-s/tk* signal in the underlying tissue of the ventromedial diencephalon.

Figure 4 illustrates various domains/motifs found in the *pan-s/tk* polypeptide.

Detailed Description of the Invention

The regulation of protein serine and threonine phosphorylation is an important mechanism for developmental control. We describe here a new class of transmembrane serine/threonine kinases, called "*pan-s/tk*" (for pancreatic serine/threonine kinases) receptors. Briefly, a cDNA fragment for a rat *pan-s/tk* gene was first isolated from a fetal pancreas cDNA library. The human EST database was queried for sequences encoding a protein related the rat *pan-s/tk* gene. A total of 60 sequences were detected. These sequences appear to be encoded by two separate genes which we refer to herein as human *pan-s/tk*-1 and human *pan-s/tk*-2.

The human *pan-s/tk*-1 gene encodes a protein of at least 520 amino acids. The available EST sequences do not contain what appears to be the initiator start codon and the protein coding sequence appears to be truncated at its amino-terminal end relative to the rat *pan-s/tk* sequence. The available 520 amino acid polypeptide is 97.6% identical to the rat *pan-s/tk* protein. There is evidence for two alternatively processed human *pan-s/tk* mRNAs. These mRNAs differ only in the length of their 3' untranslated region and both encode the same polypeptide. One sequence (human *pan-s/tk*-1A) is 1799 nt in length. Evidence for this mRNA form comes from two independent EST clones which contain a polyA at this position. The second mRNA (human pstk-1b) appears to be at least 3,063 nt in length. The additional sequence appears to contain a 3' untranslated region since there are multiple stop codons in all frames. This sequence also ends in a polyA stretch with a polyadenylation signal preceding it.

A second *pan-s/tk* like protein sequence, *pan-s/tk*-2, is encoded by a related but distinct mRNA sequence derived from seven EST sequences. This nucleotide sequence is

1,778 nt in length. The identified fragment does not appear to contain an initiator start codon and does not end in polyA. The nucleotide sequence encodes a 522 amino acid protein which is 77.6% identical to the rat *pan-s/tk* protein and 78.1% identical to the human *pan-s/tk*-1 protein.

One of the EST equences contained within the pstk-2 contig (gi#:16467796) represents the 5' sequencing read of the Image consortium clone zl81e12. The 3' read from this same clone (gi#:1646797) cannot be assembled into this 1,778 nt contig. This suggests that this 3' read represents the 3' end of the same mRNA sequence. This 3' clone was then used to constuct another contig (SEQ ID No. 8) derived from an additional 27 ESTs which does not overlap with the *pan-s/tk*-2 contig (SEQ ID No. 7) at this time. This 3' sequence ends with polyA following a polyadenylation signal, contains mutiple stop codons in all frames and exhibts no matches to known proteins (tblastn), all of which support the idea that this sequence represents the 3' untranslated region of a *pan-s/tk*-2 mRNA.

Table 1

Guide to Sequence Listing

clone	nucleic acid	polypeptide
rat pan-s/tk	SEQ ID No. 1	SEQ ID No. 2
human pan-s/tk-1A	SEQ ID No. 3	SEQ ID No. 4
human pan-s/tk-1B	SEQ ID No. 5	SEQ ID No. 6
human pan-s/tk-2	SEQ ID No. 7 SEQ ID No. 8 (3')	SEQ ID No. 9

The sequence of exemplary *pan-s/tk* genes (SEQ ID No. 1, 3, 5, 7 or 8) indicate that these genes encode a receptor-type serine/threonine kinase (SEQ ID No. 2, 4, 6 or 9) with an intracellular serine/threonine domain.

While not wishing to be bound by any particular theory, the transmembrane domain of the *pan-s/tk* protein shown in SEQ ID No. 2 is located approximately at residues 107-132. Southern blot analysis indicates two bands, a 2.2kb and 4.0kb transcript, with multiple other bands seen at lower stringency, e.g., indicating that the *pan-s/tk* gene has a number of closely related homologs or splice variants. Based on this analysis, the sequence shown in SEQ ID No. 2 is presumably from the 2.2 kb band. It is possible that this transcript encodes a membrane-anchored cytoplamic fragment (truncant) form of the protein, and the 4.0kb transcript may encode the full-length form of the protein, e.g., including an extracellular ligand binding domain.

Comparison of *pan-s/tk* sequence with other known receptor serine/threonine kinases defines a new subclass of receptor-type serine/threonine kinases. The *pan-s/tk* message was found to be srongly expressed in adult pancreas and testis, and also expressed, though at lower levels, in brain, heart and kidney. In embryos, whole mount *in situ* hybridization shows expression in developing heart, brain (e.g., forebrain), a ganglion anterior of optic vesicle, the dorsal neural tube, the pancreatic epithelium, and early gut endoderm. Expression in the gut endoderm appears to happen early and is lost as the epithelial cells mature later in development. This gene may therefore represent an early marker of endoderm epithelialization. Expression in the pancreatic epithelium increases with time in development as demonstrated by the increasing intensity of expression in the pancreas.

Despite the importance of pancreatic endocrine cells in physiology and disease, little is known at the molecular level about the developmental control of the pancreas, and no cell-cell signaling molecules have yet been identified as specific regulators of pancreatic development. We were therefore intrigued to find that *pan-s/tk* appeared specifically in the pancreas even from early organogenesis.

At the site of pancreatic development, the expression of *pan-s/tk* was first observed as early as E12, in the endodermal layer of the dorsal region of the gut while it was still open to the yolk sac. The site within the endoderm that gives rise to the pancreatic rudiment has previously been identified from morphological descriptions of early pancreatic development and from *in vitro* explant culture experiments (Wessells and Cohen, 1967). In these studies, it was found that pancreatic tissue could be cultured from a specific region of the gut from 10- and 11- somite embryos, was formed less efficiently when the tissue was obtained from 7-9 somite embryos, and was not produced from earlier embryos. The site of *pan-s/tk* expression identified in our experiments appears to include the region within the endoderm that becomes committed to form the pancreas, and moreover the time of appearance of *pan-s/tk* RNA appears to be similar to the time of initial pancreatic commitment.

Later, as the pancreatic rudiment becomes morphologically distinguishable, cells containing *pan-s/tk* appeared to be located only within the pancreatic rudiment, and not in adjacent areas of the gut, except for certain posterior regions of the intestinal loops. The temporal and spatial expression of *pan-s/tk* in developing pancreas is somewhat similar to that of STF-1, a nuclear factor that is the earliest known marker for pancreatic development (described *supra*). However, unlike STF-1, *pan-s/tk* is not expressed in the adjacent duodenum, making it a more specific marker at the site of the early developing pancreas. The *pan-s/tk* message is thus noteworthy as a particularly early and specific marker of pancreatic development.

Within the developing pancreas, *pan-s/tk* expression is not seen in all cells. We have carried out insulin staining of e18 pancreas after in situ hybridization of the *pan-s/tk* clone.

The insulin positive cells are, by and large, exclusive of those cells that express the kinase, e.g., the kinase seems to be compartmentalized in at that development stage to cells of exocrine lineage.

Accordingly, certain aspects of the present invention relate to nucleic acids encoding *pan-s/tk* polypeptides, the *pan-s/tk* polypeptides themselves (including various fragments), antibodies immunoreactive with *pan-s/tk* proteins, and preparations of such compositions. Moreover, the present invention provides diagnostic and therapeutic assays and reagents for detecting and treating disorders involving, for example, aberrant expression (or loss thereof) of *pan-s/tk*, ligands of *pan-s/tk* receptors or intracellular signal transducers thereof.

In addition, drug discovery assays are provided for identifying agents which can modulate the biological function of *pan-s/tk* proteins, such as by altering the binding of *pan-s/tk* molecules to extracellular/matrix factors or the ability of the kinase activity of the receptor to modify intracellular substrates involved in signaling from the receptor. Such agents can be useful therapeutically to alter the growth, maintenance and/or differentiation of a cell, e.g., of pancreatic, neuronal, heart or kidney tissue. Other aspects of the invention are described below or will be apparent to those skilled in the art in light of the present disclosure.

For convenience, certain terms employed in the specification, examples, and appended claims are collected here.

The term "*pan-s/tk*" refers to a family of polypeptides characterized at least in part by being identical or sharing a degree of sequence homology with all or a portion of the receptor serine/threonine kinase represented in SEQ ID No. 2, 4, 6 or 9. The *pan-s/tk* polypeptides can be cloned or purified from any of a number of eukaryotic organisms, especially vertebrates, and particularly mammals. Moreover, other *pan-s/tk* polypeptides can be generated according to the present invention, which polypeptides do not ordinarily exist in nature, but rather are generated by non-natural mutagenic techniques.

A "glycosylated" *pan-s/tk* polypeptide is an *pan-s/tk* polypeptide having a covalent linkage with a glycosyl group (e.g. a derivatized with a carbohydrate). For instance, the *pan-s/tk* protein can be glycosylated on an existing residue, or can be mutated to preclude carbohydrate attachment, or can be mutated to provide new glycosylation sites, such as for N-linked or O-linked glycosylation.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid comprising an open reading frame encoding of a *pan-s/tk* polypeptide, including both exon and (optionally) intron sequences. A "recombinant gene" refers to nucleic acid encoding a *pan-s/tk* polypeptide and comprising *pan-s/tk*-encoding exon sequences, though it may optionally include intron sequences which are derived from, for example, a chromosomal *pan-s/tk* gene or from an unrelated chromosomal gene. Exemplary recombinant genes encoding the subject *pan-s/tk* polypeptide are represented in the appended Sequence Listing. The term "intron" refers to a DNA sequence present in a given *pan-s/tk* gene which is not translated into protein and is generally found between exons.

As used herein, the term "transfection" means the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a recombinant form of a *pan-s/tk* polypeptide or, where anti-sense expression occurs from the transferred gene, the expression of a naturally-occurring form of the *pan-s/tk* protein is disrupted.

As used herein, the term "specifically hybridizes" refers to the ability of a nucleic acid probe/primer of the invention to hybridize to at least 15 consecutive nucleotides of a *pan-s/tk* gene, such as a *pan-s/tk* sequence designated in SEQ ID No. 1, 3, 5, 7 or 8, or a sequence complementary thereto, or naturally occurring mutants thereof, such that it has less than 15%, preferably less than 10%, and more preferably less than 5% background hybridization to a cellular nucleic acid (e.g., mRNA or genomic DNA) encoding a protein other than a *pan-s/tk* protein, as defined herein. In preferred embodiments, the oligonucleotide probe specifically detects only a *pan-s/tk* gene, e.g., it does not substantially hybridize to transcripts for other RSTKs, such as the TGF-β or activin receptors ALK-1-7.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

"Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, which induce or control transcription of protein coding sequences with which they are operably linked. In preferred embodiments, transcription of a recombinant *pan-s/tk* gene is under the control of a promoter sequence (or other transcriptional regulatory sequence) which controls the expression of the recombinant gene in a cell-type in which expression is intended. It will also be understood that the recombinant gene can be under the control of transcriptional regulatory sequences which are the same or which are different from those sequences which control transcription of the naturally-occurring forms of *pan-s/tk* genes.

As used herein, the term "tissue-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue, such as cells of neuronal or hematopoietic origin. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but can cause at least low level expression in other tissues as well.

As used herein, a "transgenic animal" is any animal, preferably a non-human mammal, bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In an exemplary transgenic animal, the transgene causes cells to express a recombinant form of a *pan-s/tk* protein, e.g. either agonistic or antagonistic forms. However, transgenic animals in which the recombinant *pan-s/tk* gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. Moreover, "transgenic animal" also includes those recombinant animals in which gene disruption of one or more *pan-s/tk* genes is caused by human intervention, including both recombination and antisense techniques.

The "non-human animals" of the invention include vertebrates such as rodents, non-human primates, livestock, avian species, amphibians, reptiles, etc. The term "chimeric animal" is used herein to refer to animals in which the recombinant gene is found, or in which the recombinant is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that a recombinant *pan-s/tk* genes is present and/or expressed or disrupted in some tissues but not others.

As used herein, the term "transgene" means a nucleic acid sequence (encoding, e.g., a *pan-s/tk* polypeptide, or pending an antisense transcript thereto), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout of the endogenous *pan-s/tk* gene). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid.

As is well known, genes for a particular polypeptide may exist in single or multiple copies within the genome of an individual. Such duplicate genes may be identical or may have certain modifications, including nucleotide substitutions, additions or deletions, which all still code for polypeptides having substantially the same activity. The term "DNA sequence encoding a *pan-s/tk* polypeptide" may thus refer to one or more genes within a particular individual. Moreover, certain differences in nucleotide sequences may exist between individuals of the same species, which are called alleles. Such allelic differences may or may not result in differences in amino acid sequence of the encoded polypeptide yet still encode a protein with the same biological activity.

"Homology" and "identity" each refer to sequence similarity between two polypeptide sequences, with identity being a more strict comparison. Homology and identity can each be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same amino acid residue, then the polypeptides can be referred to as identical at that position; when the equivalent site is occupied by the same amino acid (e.g., identical) or a similar amino acid (e.g., similar in steric and/or electronic nature), then the molecules can be refered to as homologous at that position. A percentage of homology or identity between sequences is a function of the number of matching or homologous positions shared by the sequences. An "unrelated" or "non-homologous" sequence shares less than 40 percent identity, though preferably less than 25 percent identity, with a *pan-s/tk* sequence of the present invention.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A "chimeric protein" or "fusion protein" is a fusion of a first amino acid sequence encoding a *pan-s/tk* polypeptide with a second amino acid sequence defining a domain (e.g. polypeptide portion) foreign to and not substantially homologous with any domain of a naturally-occuring *pan-s/tk* protein. A chimeric protein may present a foreign domain which is found (albeit in a different protein) in an organism which also expresses the first protein, or it may be an "interspecies", "intergenic", etc. fusion of protein structures expressed by different kinds of organisms. In general, a fusion protein can be represented by the general formula $(X)_n$-$(Y)_m$-$(Z)_n$, wherein Y represents all or a portion of the *pan-s/tk* protein, X and Z each independently represent an amino acid sequences which are not naturally found as a polypeptide chain contiguous with the *pan-s/tk* sequence, m is an integer greater than or equal to 1, and each occurrence of n is, independently, 0 or an integer greater than or equal to 1 (n and m are preferably no greater than 5 or 10).

The term "isolated" as also used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs, or RNAs, respectively, that are present in the natural source of the macromolecule. For example, an isolated nucleic acid encoding a *pan-s/tk* polypeptide preferably includes no more than 10 kilobases (kb) of nucleic acid sequence which naturally immediately flanks the *pan-s/tk* gene in genomic DNA, more preferably no more than 5kb of such naturally occurring flanking sequences, and most preferably less than 1.5kb of such naturally occurring flanking sequence. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state.

The term "ligand" refers to any protein or proteins that can interact with the *pan-r/tk* receptor ligand binding domain. The ligand or ligands can be soluble or membrane bound. The ligand or ligands can be a naturally occurring protein, or synthetically or recombinantly produced. The ligand can also be a nonprotein molecule that acts as ligand when it interacts with the *pan-r/tk* receptor binding domain. Interactions between the ligand and receptor binding domain include, but are not limited to, any covalent or non-covalent interactions. The receptor binding domain is any region (extracellular) of the *pan-r/tk* receptor molecule that interacts directly or indirectly with the *pan-r/tk* ligand. Agonists and antagonists of *pan-r/tk* that can interact with the *pan-r/tk* receptor binding domain are ligands.

As described below, one aspect of the invention pertains to isolated nucleic acids comprising nucleotide sequences encoding *pan-s/tk* polypeptides, and/or equivalents of such nucleic acids. The term nucleic acid as used herein is intended to include fragments as equivalents. The term equivalent is understood to include nucleotide sequences encoding functionally equivalent *pan-s/tk* polypeptides or functionally equivalent peptides having an

activity of a *pan-s/tk* protein such as described herein. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will, therefore, include sequences that differ from the nucleotide sequence of the *pan-s/tk* coding sequence shown in SEQ ID No. 1, 3, 5, 7 or 8 due to the degeneracy of the genetic code. Equivalents will also include nucleotide sequences that hybridize under stringent conditions (i.e., equivalent to about 20-27°C below the melting temperature (T_m) of the DNA duplex formed in about 1M salt) to the nucleotide sequences represented in SEQ ID No. 1, 3, 5, 7 or 8. In one embodiment, equivalents will further include nucleic acid sequences derived from and evolutionarily related to, a nucleotide sequences shown in SEQ ID No. 1, 3, 5, 7 or 8.

Moreover, it will be generally appreciated that, under certain circumstances, it may be advantageous to provide homologs of a *pan-s/tk* polypeptide which function in a limited capacity as one of either a *pan-s/tk* agonist (mimetic) or a *pan-s/tk* antagonist, in order to promote or inhibit only a subset of the biological activities of the naturally-occurring form of the protein. Thus, specific biological effects can be elicited by treatment with a homolog of limited function. For example, truncated forms of the receptor, e.g., soluble fragments of the extracellular domain, can be provided to competitively inhibit ligand binding to the receptor. Likewise, mutants having altered kinase activity profiles, e.g., altered k_{cat} or k_m or constitutively active mutants, can be provided.

Homologs of the subject *pan-s/tk* protein can be generated by mutagenesis, such as by discrete point mutation(s), or by truncation. For instance, mutation can give rise to homologs which retain substantially the same, or merely a subset, of the biological activity of the *pan-s/tk* polypeptide from which it was derived. Alternatively, antagonistic forms of the protein can be generated which are able to inhibit the function of the naturally occurring form of the protein, such as by competitively binding to a *pan-s/tk* substrate or *pan-s/tk* associated protein, as for example competing with wild-type *pan-s/tk* in the binding of an extracellular ligand, or binding to an intracellular protein such as a substrate of the kinase activity. Thus, the *pan-s/tk* protein and homologs thereof provided by the subject invention may be either positive or negative regulators of cell growth, death and/or differentiation.

In general, polypeptides referred to herein as having an activity of a *pan-s/tk* protein (e.g., are "bioactive") are defined as polypeptides which include an amino acid sequence corresponding (e.g., identical or homologous) to all or a portion of the amino acid sequences of the *pan-s/tk* protein shown in SEQ ID No. 2, 4, 6 or 9, and which mimic or antagonize all or a portion of the biological/biochemical activities of a naturally occurring *pan-s/tk* protein. Examples of such biological activity include: the ability to phosphorylate a serine or threonine residue. The bioactivity of certain embodiments of the *pan-s/tk* protein can be characterized in terms of an ability to regulate differentiation and/or maintenance of pancreatic and neural cells and tissue.

Other biological activities of the subject *pan-s/tk* proteins are described herein or will be reasonably apparent to those skilled in the art. According to the present invention, a polypeptide has biological activity if it is a specific agonist or antagonist of a naturally-occurring form of a *pan-s/tk* protein.

Preferred nucleic acids encode a *pan-s/tk* polypeptide comprising an amino acid sequence at least 60%, 70% or 80% homologous, more preferably at least 85% homologous and most preferably at least 95% homologous with an amino acid sequence of a naturally occurring *pan-s/tk* protein, e.g., such as represented in SEQ ID No. 2, 4, 6 or 9. Nucleic acids which encode polypeptides at least about 98-99% homology with an amino acid sequence represented in SEQ ID No. 2, 4, 6 or 9 are of course also within the scope of the invention, as are nucleic acids identical in sequence with the enumerated *pan-s/tk* sequence of the sequence listing. In one embodiment, the nucleic acid is a cDNA encoding a polypeptide having at least one activity of the subject *pan-s/tk* polypeptide.

In certain preferred embodiments, the invention features a purified or recombinant *pan-s/tk* polypeptide. It will be understood that certain post-translational modifications, e.g., glycosylation, phosphorylation and the like, can increase the apparent molecular weight of the *pan-s/tk* protein relative to the unmodified polypeptide chain, and cleavage of certain sequences, such as pro-sequences, can likewise decrease the apparent molecular weight. Other preferred *pan-s/tk* polypeptides include a mature, extracellular fragment (soluble) of the receptor. Yet other preferred *pan-s/tk* polypeptides include an intracellular domain, e.g., including the serine/threonine kinase domain. Another aspect of the invention provides a nucleic acid which hybridizes under high or low stringency conditions to the nucleic acid represented by SEQ ID No. 1, 3, 5, 7 or 8. Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C.

Nucleic acids, having a sequence that differs from the nucleotide sequences shown in SEQ ID No. 1, 3, 5, 7 or 8 due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent peptides (i.e., a peptide having a biological activity of a *pan-s/tk* polypeptide) but differ in sequence from the sequence shown in the sequence listing due to degeneracy in the genetic code. For example, a number of amino acids are designated by more than one triplet. Codons that specify the same amino acid, or synonyms (for example, CAU and CAC each encode histidine) may

result in "silent" mutations which do not affect the amino acid sequence of a *pan-s/tk* polypeptide. However, it is expected that DNA sequence polymorphisms that do lead to changes in the amino acid sequences of the subject *pan-s/tk* polypeptides will exist among, for example, humans. One skilled in the art will appreciate that these variations in one or more nucleotides (up to about 3-5% of the nucleotides) of the nucleic acids encoding polypeptides having an activity of a *pan-s/tk* polypeptide may exist among individuals of a given species due to natural allelic variation.

As used herein, a *pan-s/tk* gene fragment refers to a nucleic acid having fewer nucleotides than the nucleotide sequence encoding the entire mature form of a *pan-s/tk* protein yet which (preferably) encodes a polypeptide which retains some biological activity of the full length protein. Fragment sizes contemplated by the present invention include, for example, 5, 10, 25, 50, 75, 100, or 200 (contiguous) amino acids in length. In a preferred embodiment of a truncated receptor, the polypeptide will include all or a sufficient portion of the extracellular domain to bind to a *pan-s/tk* ligand. In another, the polypeptide includes the kinase domain of the cytosolic portion of the protein. In either embodiment, the *pan-s/tk* polypeptide can also include the transmembrane domain, particularly where membrane localized (instead of soluble) fragments of the *pan-s/tk* protein are desired.

As indicated by the examples set out below, *pan-s/tk* protein-encoding nucleic acids can be obtained from mRNA present in cells of metazoan organisms. It should also be possible to obtain nucleic acids encoding *pan-s/tk* polypeptides of the present invention from genomic DNA from both adults and embryos. For example, a gene encoding a *pan-s/tk* protein can be cloned from either a cDNA or a genomic library in accordance with protocols described herein, as well as those generally known to persons skilled in the art. A cDNA encoding a *pan-s/tk* protein can be obtained by isolating total mRNA from a cell, such as a mammalian cell, e.g. a human cell, as desired. Double stranded cDNAs can be prepared from the total mRNA, and subsequently inserted into a suitable plasmid or bacteriophage vector using any one of a number of known techniques. The gene encoding a *pan-s/tk* protein can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acid of the invention can be DNA or RNA. A preferred nucleic acid is a cDNA including a nucleotide sequence represented by one of SEQ ID No. 1, 3, 5, 7 or 8.

Another aspect of the invention relates to the use of the isolated nucleic acid in "antisense" therapy. As used herein, "antisense" therapy refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridize (e.g. binds) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding a subject *pan-s/tk* protein so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific

interactions in the major groove of the double helix. In general, "antisense" therapy refers to the range of techniques generally employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes a *pan-s/tk* protein. Alternatively, the antisense construct is an oligonucleotide probe which is generated *ex vivo* and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences of a *pan-s/tk* gene. Such oligonucleotide probes are preferably modified oligonucleotides which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and are therefore stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patents 5,176,996; 5,264,564; and 5,256,775), or peptide nucleic acids (PNAs). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by Van der Krol et al. (1988) Biotechniques 6:958-976; and Stein et al. (1988) Cancer Res 48:2659-2668.

Accordingly, the modified oligomers of the invention are useful in therapeutic, diagnostic, and research contexts. In therapeutic applications, the oligomers are utilized in a manner appropriate for antisense therapy in general. For such therapy, the oligomers of the invention can be formulated for a variety of routes of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remmington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For injection, the oligomers of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the oligomers may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

Systemic administration can also be by transmucosal or transdermal means, or the compounds can be administered orally. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For oral administration, the oligomers are formulated into conventional oral administration forms such as capsules, tablets, and tonics. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art.

In addition to use in therapy, the oligomers of the invention may be used as diagnostic reagents to detect the presence or absence of the target DNA or RNA sequences to which they specifically bind. Such diagnostic tests are described in further detail below.

Likewise, the antisense constructs of the present invention, by antagonizing the normal biological activity of a *pan-s/tk* protein, e.g., by reducing the level of its expression, can be used in the manipulation of tissue, e.g. tissue maintenance, differentiation or growth, both *in vivo* and *ex vivo*.

Furthermore, the anti-sense techniques (e.g. microinjection of antisense molecules, or transfection with plasmids whose transcripts are anti-sense with regard to a *pan-s/tk* mRNA or gene sequence) can be used to investigate the role of *pan-s/tk* in developmental events, as well as the normal cellular function of *pan-s/tk* in adult tissue. Such techniques can be utilized in cell culture, but can also be used in the creation of transgenic animals (described *infra*).

This invention also provides expression vectors containing a nucleic acid encoding a *pan-s/tk* polypeptide, operably linked to at least one transcriptional regulatory sequence. Operably linked is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Regulatory sequences are art-recognized and are selected to direct expression of the subject *pan-s/tk* proteins. Accordingly, the term transcriptional regulatory sequence includes promoters, enhancers and other expression control elements. Such regulatory sequences are described in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). For instance, any of a wide variety of expression control sequences, sequences that control the expression of a DNA sequence when operatively linked to it, may be used in these vectors to express DNA sequences encoding *pan-s/tk* polypeptides of this invention. Such useful expression control sequences, include, for example, a viral LTR, such as the LTR of the Moloney murine leukemia virus, the early and late promoters of SV40, adenovirus or cytomegalovirus immediate early promoter, the lac system, the trp system, the TAC or TRC system, T7 promoter whose expression is directed by T7 RNA polymerase, the major operator and promoter regions of phage λ, the control regions for fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast α-mating factors, the polyhedron promoter of the baculovirus system and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed.

Moreover, the vector's copy number, the ability to control that copy number and the expression of any other proteins encoded by the vector, such as antibiotic markers, should

also be considered. In one embodiment, the expression vector includes a recombinant gene encoding a polypeptide having an agonistic activity of a subject *pan-s/tk* polypeptide, such as a constitutively active intracellular kinase domain, or alternatively, encoding a polypeptide which is an antagonistic form of the *pan-s/tk* protein, such as a soluble truncated form including the extracellular ligand binding domain. Such expression vectors can be used to transfect cells and thereby produce polypeptides, including fusion proteins, encoded by nucleic acids as described herein.

Moreover, the gene constructs of the present invention can also be used as a part of a gene therapy protocol to deliver nucleic acids, e.g., encoding either an agonistic or antagonistic form of a subject *pan-s/tk* proteins or an antisense molecule described above. Thus, another aspect of the invention features expression vectors for *in vivo* or *in vitro* transfection and expression of a *pan-s/tk* polypeptide or antisense molecule in particular cell types so as to reconstitute the function of, or alternatively, abrogate all or a portion of the biological function of *pan-s/tk*-induced transcription in a tissue in which the naturally-occurring form of the protein is misexpressed (or has been disrupted); or to deliver a form of the protein which alters maintenance or differentiation of tissue, or which inhibits neoplastic or hyperplastic proliferation.

Expression constructs of the subject *pan-s/tk* polypeptides, as well as antisense constructs, may be administered in any biologically effective carrier, e.g. any formulation or composition capable of effectively delivering the recombinant gene to cells *in vivo*. Approaches include insertion of the subject gene in viral vectors including recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1, or recombinant bacterial or eukaryotic plasmids. Viral vectors transfect cells directly; plasmid DNA can be delivered with the help of, for example, cationic liposomes (lipofectin) or derivatized (e.g. antibody conjugated), polylysine conjugates, gramacidin S, artificial viral envelopes or other such intracellular carriers, as well as direct injection of the gene construct or $CaPO_4$ precipitation carried out *in vivo*. It will be appreciated that because transduction of appropriate target cells represents the critical first step in gene therapy, choice of the particular gene delivery system will depend on such factors as the phenotype of the intended target and the route of administration, e.g. locally or systemically. Furthermore, it will be recognized that the particular gene construct provided for *in vivo* transduction of *pan-s/tk* expression are also useful for *in vitro* transduction of cells, such as for use in the *ex vivo* tissue culture systems described below.

A preferred approach for *in vivo* introduction of nucleic acid into a cell is by use of a viral vector containing nucleic acid, e.g. a cDNA encoding the particular *pan-s/tk* polypeptide desired. Infection of cells with a viral vector has the advantage that a large proportion of the targeted cells can receive the nucleic acid. Additionally, molecules encoded within the viral vector, e.g., by a cDNA contained in the viral vector, are expressed

efficiently in cells which have taken up viral vector nucleic acid. Retrovirus vectors, adenovirus vectors and adeno-associated virus vectors are exemplary recombinant gene delivery system for the transfer of exogenous genes *in vivo*, particularly into humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host.

In addition to viral transfer methods, such as those illustrated above, non-viral methods can also be employed to cause expression of a subject *pan-s/tk* polypeptide in the tissue of an animal. Most nonviral methods of gene transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral gene delivery systems of the present invention rely on endocytic pathways for the uptake of the subject *pan-s/tk* polypeptide gene by the targeted cell. Exemplary gene delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

In clinical settings, the gene delivery systems for the therapeutic *pan-s/tk* gene can be introduced into a patient-animal by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g. by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (see U.S. Patent 5,328,470) or by stereotactic injection (e.g. Chen et al. (1994) PNAS 91: 3054-3057). A *pan-s/tk* gene can be delivered in a gene therapy construct by electroporation using techniques described, for example, by Dev et al. ((1994) Cancer Treat Rev 20:105-115).

The pharmaceutical preparation of the gene therapy construct can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery system can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system.

Another aspect of the present invention concerns recombinant forms of the *pan-s/tk* proteins. Recombinant polypeptides preferred by the present invention, in addition to native *pan-s/tk* proteins, are at least 60% or 70% homologous, more preferably at least 80% homologous and most preferably at least 85% homologous with an amino acid sequence represented by SEQ ID No. 2, 4, 6 or 9. Polypeptides which possess an activity of a *pan-s/tk* protein (i.e. either agonistic or antagonistic), and which are at least 90%, more

preferably at least 95%, and most preferably at least about 98-99% homologous with SEQ ID No. 2, 4, 6 or 9 are also within the scope of the invention. Such polypeptides, as described above, include various truncated forms of the protein.

The term "recombinant *pan-s/tk* polypeptide" refers to a polypeptide which is produced by recombinant DNA techniques, wherein generally, DNA encoding a *pan-s/tk* polypeptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein. Moreover, the phrase "derived from", with respect to a recombinant *pan-s/tk* gene, is meant to include within the meaning of "recombinant protein" those proteins having an amino acid sequence of a native *pan-s/tk* protein, or an amino acid sequence similar thereto which is generated by mutations including substitutions and deletions (including truncation) of a naturally occurring form of the protein.

The present invention further pertains to recombinant forms of the subject *pan-s/tk* polypeptides which are encoded by genes derived from a mammal (e.g. a human), bird, reptile or amphibian and which have amino acid sequences evolutionarily related to the *pan-s/tk* protein represented in SEQ ID No. 2, 4, 6 or 9. Such recombinant *pan-s/tk* polypeptides preferably are capable of functioning in one of either role of an agonist or antagonist of at least one biological activity of a wild-type ("authentic") *pan-s/tk* protein of the appended sequence listing. The term "evolutionarily related to", with respect to amino acid sequences of *pan-s/tk* proteins, refers to both polypeptides having amino acid sequences which have arisen naturally, and also to mutational variants of *pan-s/tk* polypeptides which are derived, for example, by combinatorial mutagenesis.

The present invention also provides methods of producing the subject *pan-s/tk* polypeptides. For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding the subject polypeptides can be cultured under appropriate conditions to allow expression of the peptide to occur. The cells may be harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art. The recombinant *pan-s/tk* polypeptide can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such peptide. In a preferred embodiment, particularly for version of the subject polypeptide which do not include the transmembrabe domain (such as truncated extracellular and intracellular fragments, the recombinant *pan-s/tk* polypeptide is a fusion protein containing a domain which facilitates its purification, such as GST fusion protein or poly(His) fusion protein.

This invention also pertains to a host cell transfected to express recombinant forms of the subject *pan-s/tk* polypeptides. The host cell may be any eukaryotic or prokaryotic

cell. Thus, a nucleotide sequence derived from the cloning of *pan-s/tk* proteins, encoding all or a selected portion of a full-length protein, can be used to produce a recombinant form of a *pan-s/tk* polypeptide via microbial or eukaryotic cellular processes. Ligating the polynucleotide sequence into a gene construct, such as an expression vector, and transforming or transfecting into hosts, either eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial cells), are standard procedures used in producing other well-known proteins, e.g. MEKK, Src, and a wide range of receptors kinases, and the like. Similar procedures, or modifications thereof, can be employed to prepare recombinant *pan-s/tk* polypeptides by microbial means or tissue-culture technology in accord with the subject invention.

The recombinant *pan-s/tk* genes can be produced by ligating nucleic acid encoding an *pan-s/tk* protein, or a portion thereof, into a vector suitable for expression in either prokaryotic cells, eukaryotic cells, or both. Expression vectors for production of recombinant forms of the subject *pan-s/tk* polypeptides include plasmids and other vectors. For instance, suitable vectors for the expression of a *pan-s/tk* polypeptide include plasmids of the types: pBR322-derived plasmids, pEMBL-derived plasmids, pEX-derived plasmids, pBTac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as E. coli.

A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into S. cerevisiae (see, for example, Broach et al. (1983) in Experimental Manipulation of Gene Expression, ed. M. Inouye Academic Press, p. 83, incorporated by reference herein). These vectors can replicate in E. coli due the presence of the pBR322 ori, and in S. cerevisiae due to the replication determinant of the yeast 2 micron plasmid. In addition, drug resistance markers such as ampicillin can be used. In an illustrative embodiment, a *pan-s/tk* polypeptide is produced recombinantly utilizing an expression vector generated by sub-cloning the coding sequence of a *pan-s/tk* gene represented in SEQ ID No. 1, 3, 5, 7 or 8.

The preferred mammalian expression vectors contain both prokaryotic sequences, to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The pcDNAI/amp, pcDNAI/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the bovine papillomavirus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells. The various methods employed

in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989) Chapters 16 and 17.

In some instances, it may be desirable to express the recombinant *pan-s/tk* polypeptide by the use of a baculovirus expression system. Examples of such baculovirus expression systems include pVL-derived vectors (such as pVL1392, pVL1393 and pVL941), pAcUW-derived vectors (such as pAcUW1), and pBlueBac-derived vectors (such as the ß-gal containing pBlueBac III).

When it is desirable to express only a portion of a *pan-s/tk* protein, such as a form lacking a portion of the N-terminus, i.e. a truncation mutant which lacks the signal peptide, it may be necessary to add a start codon (ATG) to the oligonucleotide fragment containing the desired sequence to be expressed. It is well known in the art that a methionine at the N-terminal position can be enzymatically cleaved by the use of the enzyme methionine aminopeptidase (MAP). MAP has been cloned from E. coli (Ben-Bassat et al. (1987) J. Bacteriol. 169:751-757) and Salmonella typhimurium and its *in vitro* activity has been demonstrated on recombinant proteins (Miller et al. (1987) PNAS 84:2718-1722). Therefore, removal of an N-terminal methionine, if desired, can be achieved either *in vivo* by expressing *pan-s/tk*-derived polypeptides in a host which produces MAP (e.g., E. coli or CM89 or S. cerevisiae), or *in vitro* by use of purified MAP (e.g., procedure of Miller et al., supra).

Alternatively, the coding sequences for the polypeptide can be incorporated as a part of a fusion gene including a nucleotide sequence encoding a different polypeptide. This type of expression system can be useful under conditions where it is desirable to produce an immunogenic fragment of a *pan-s/tk* protein. For example, the VP6 capsid protein of rotavirus can be used as an immunologic carrier protein for portions of the *pan-s/tk* polypeptide, either in the monomeric form or in the form of a viral particle. The nucleic acid sequences corresponding to the portion of a subject *pan-s/tk* protein to which antibodies are to be raised can be incorporated into a fusion gene construct which includes coding sequences for a late vaccinia virus structural protein to produce a set of recombinant viruses expressing fusion proteins comprising *pan-s/tk* epitopes as part of the virion. It has been demonstrated with the use of immunogenic fusion proteins utilizing the Hepatitis B surface antigen fusion proteins that recombinant Hepatitis B virions can be utilized in this role as well. Similarly, chimeric constructs coding for fusion proteins containing a portion of a *pan-s/tk* protein and the poliovirus capsid protein can be created to enhance immunogenicity of the set of polypeptide antigens (see, for example, EP Publication No:

0259149; and Evans et al. (1989) Nature 339:385; Huang et al. (1988) J. Virol. 62:3855; and Schlienger et al. (1992) J. Virol. 66:2).

The Multiple Antigen Peptide system for peptide-based immunization can also be utilized to generate an immunogen, wherein a desired portion of a *pan-s/tk* polypeptide is obtained directly from organo-chemical synthesis of the peptide onto an oligomeric branching lysine core (see, for example, Posnett et al. (1988) JBC 263:1719 and Nardelli et al. (1992) J. Immunol. 148:914). Antigenic determinants of *pan-s/tk* proteins can also be expressed and presented by bacterial cells.

In addition to utilizing fusion proteins to enhance immunogenicity, it is widely appreciated that fusion proteins can also facilitate the expression of proteins, and accordingly, can be used in the expression of the *pan-s/tk* polypeptides of the present invention, particularly truncated forms of the *pan-s/tk* protein. For example, *pan-s/tk* polypeptides can be generated as glutathione-S-transferase (GST-fusion) proteins. Such GST-fusion proteins can enable easy purification of the *pan-s/tk* polypeptide, as for example by the use of glutathione-derivatized matrices (see, for example, Current Protocols in Molecular Biology, eds. Ausubel et al. (N.Y.: John Wiley & Sons, 1991)).

In another embodiment, a fusion gene coding for a purification leader sequence, such as a poly-(His)/enterokinase cleavage site sequence at the N-terminus of the desired portion of the recombinant protein, can allow purification of the expressed fusion protein by affinity chromatography using a Ni2+ metal resin. The purification leader sequence can then be subsequently removed by treatment with enterokinase to provide the purified protein (e.g., see Hochuli et al. (1987) J. Chromatography 411:177; and Janknecht et al. PNAS 88:8972).

Techniques for making fusion genes are known to those skilled in the art. Essentially, the joining of various DNA fragments coding for different polypeptide sequences is performed in accordance with conventional techniques, employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed to generate a chimeric gene sequence (see, for example, Current Protocols in Molecular Biology, eds. Ausubel et al. John Wiley & Sons: 1992).

The *pan-s/tk* polypeptides may also be chemically modified to create *pan-s/tk* derivatives by forming covalent or aggregate conjugates with other chemical moieties, such

as glycosyl groups, lipids, phosphate, acetyl groups and the like. Covalent derivatives of *pan-s/tk* proteins can be prepared by linking the chemical moieties to functional groups on amino acid sidechains of the protein or at the N-terminus or at the C-terminus of the polypeptide.

As appropriate, formulations of multimeric *pan-s/tk* receptors are also provided. The multimers of the soluble forms of the subject *pan-s/tk* receptors may be produced according to the methods known in the art. In one embodiment, the *pan-s/tk* multimers are cross-linked chemically by using known methods which will result in the formation of either dimers or higher multimers of the soluble forms of the *pan-s/tk* receptor. Another way of producing the multimers of the soluble forms of the *pan-s/tk* receptor is by recombinant techniques, e.g., by inclusion of hinge regions. This linker can facilitate enhanced flexibility of the chimeric protein allowing the various *pan-s/tk* monomeric subunits to freely and (optionally) simultaneously interact with a *pan-s/tk* ligand by reducing steric hindrance between the two fragments, as well as allowing appropriate folding of each portion to occur. The linker can be of natural origin, such as a sequence determined to exist in random coil between two domains of a protein. Alternatively, the linker can be of synthetic origin. For instance, the sequence $(Gly_4Ser)_3$ can be used as a synthetic unstructured linker. Linkers of this type are described in Huston et al. (1988) PNAS 85:4879; and U.S. Patent Nos. 5,091,513 and 5,258,498. Naturally occurring unstructured linkers of human origin are preferred as they reduce the risk of immunogenicity.

Each multimer comprises two or more monomers, each comprising the soluble form of a *pan-s/tk* receptor or a salt or functional derivative thereof. The upper limit for the number of monomers in a multimer is not important and liposomes having many such monomers thereon may be used. Such multimers preferably have 2-5 monomers and more preferably 2 or 3.

The present invention also makes available isolated *pan-s/tk* polypeptides which are isolated from, or otherwise substantially free of other cellular proteins, especially other signal transduction factors membrane-localized proteins which may normally be associated with the *pan-s/tk* polypeptide. The term "substantially free of other cellular proteins" (also referred to herein as "contaminating proteins") or "substantially pure or purified preparations" are defined as encompassing preparations of *pan-s/tk* polypeptides having less than 20% (by dry weight) contaminating protein, and preferably having less than 5% contaminating protein. Functional forms of the subject polypeptides can be prepared, for the first time, as purified preparations by using a cloned gene as described herein. By "purified", it is meant, when referring to a peptide or DNA or RNA sequence, that the indicated molecule is present in the substantial absence of other biological macromolecules, such as other proteins. The term "purified" as used herein preferably means at least 80% by

dry weight, more preferably in the range of 95-99% by weight, and most preferably at least 99.8% by weight, of biological macromolecules of the same type present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present). The term "pure" as used herein preferably has the same numerical limits as "purified" immediately above. "Isolated" and "purified" do not encompass either natural materials in their native state or natural materials that have been separated into components (e.g., in an acrylamide gel) but not obtained either as pure (e.g. lacking contaminating proteins, or chromatography reagents such as denaturing agents and polymers, e.g. acrylamide or agarose) substances or solutions. In preferred embodiments, purified *pan-s/tk* preparations will lack any contaminating proteins from the same animal from that *pan-s/tk* is normally produced, as can be accomplished by recombinant expression of, for example, a mammalian *pan-s/tk* protein in a yeast or bacterial cell.

As described above for recombinant polypeptides, isolated *pan-s/tk* polypeptides can include all or a portion of an amino acid sequences corresponding to a *pan-s/tk* polypeptide represented in SEQ ID No. 2, 4, 6 or 9 or homologous sequences thereto.

Isolated peptidyl portions of *pan-s/tk* proteins can also be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, a *pan-s/tk* polypeptide of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of a wild-type (e.g., "authentic") *pan-s/tk* protein. For example, Román et al. (1994) *Eur J Biochem* 222:65-73 describe the use of competitive-binding assays using short, overlapping synthetic peptides from larger proteins to identify binding domains.

The recombinant *pan-s/tk* polypeptides of the present invention also include homologs of the authentic *pan-s/tk* proteins, such as versions of those protein which are resistant to proteolytic cleavage, as for example, due to mutations which alter ubiquitination, enzymatic release of the extracellular domain, or other enzymatic targeting associated with the protein.

Modification of the structure of the subject *pan-s/tk* polypeptides can be for such purposes as enhancing therapeutic or prophylactic efficacy, stability (e.g., *ex vivo* shelf life and resistance to proteolytic degradation *in vivo*), or post-translational modifications (e.g., to alter glycosylation or phosphorylation patterns of protein). Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, or to produce specific antagonists thereof, are considered functional equivalents of the *pan-*

s/tk polypeptides (though they may be agonistic or antagonistic of the bioactivities of the authentic protein). Such modified peptides can be produced, for instance, by amino acid substitution, deletion, or addition.

For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e. isosteric and/or isoelectric mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur - containing = cysteine and methionine. (see, for example, Biochemistry, 2nd ed., Ed. by L. Stryer, WH Freeman and Co.: 1981). Whether a change in the amino acid sequence of a peptide results in a functional *pan-s/tk* homolog (e.g. functional in the sense that the resulting polypeptide mimics or antagonizes the authentic form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type protein, or competitively inhibit such a response. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

This invention further contemplates a method for generating sets of combinatorial point mutants of the subject *pan-s/tk* proteins as well as truncation mutants, and is especially useful for identifying potential variant sequences (e.g. homologs) that are functional in modulating signal transduction and/or ligand binding. The purpose of screening such combinatorial libraries is to generate, for example, novel *pan-s/tk* homologs which can act as either agonists or antagonist, or alternatively, possess novel activities all together. To illustrate, *pan-s/tk* homologs can be engineered by the present method to provide selective, constitutive activation of kinase activity, or alternatively, to be dominant negative inhibitors of *pan-s/tk*-dependent signal transduction. For instance, mutagenesis can provide *pan-s/tk* homologs which are able to bind extracellular ligands yet be unable to bind or signal through intracellular regulatory proteins.

In one aspect of this method, the amino acid sequences for a population of *pan-s/tk* homologs or other related proteins are aligned, preferably to promote the highest homology

possible. Such a population of variants can include, for example, *pan-s/tk* homologs from one or more species. Amino acids which appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences. In a preferred embodiment, the variegated library of *pan-s/tk* variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For instance, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential *pan-s/tk* sequences are expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g. for phage display) containing the set of *pan-s/tk* sequences therein.

There are many ways by which such libraries of potential *pan-s/tk* homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential *pan-s/tk* sequences. The synthesis of degenerate oligonucleotides is well known in the art (see for example, Narang, SA (1983) Tetrahedron 39:3; Itakura et al. (1981) Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) Science 249:386-390; Roberts et al. (1992) PNAS 89:2429-2433; Devlin et al. (1990) Science 249: 404-406; Cwirla et al. (1990) PNAS 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

Likewise, a library of coding sequence fragments can be provided for a *pan-s/tk* clone in order to generate a variegated population of *pan-s/tk* fragments for screening and subsequent selection of bioactive fragments. A variety of techniques are known in the art for generating such libraries, including chemical synthesis. In one embodiment, a library of coding sequence fragments can be generated by (i) treating a double stranded PCR fragment of a *pan-s/tk* coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule; (ii) denaturing the double stranded DNA; (iii) renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products; (iv) removing single stranded portions from reformed duplexes by treatment with S1 nuclease; and (v) ligating the resulting fragment library into an expression vector. By this exemplary method, an expression library can be derived which codes for N-terminal, C-terminal and internal fragments of various sizes.

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a certain property. Such techniques will be generally

adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of *pan-s/tk* homologs. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected.

In an exemplary embodiment, a library of variants derived from a truncated extracellular domain which are mutated, e.g., by alanine scanning mutagenesis, is expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, a large number of phage can be screened at one time. Second, since each infectious phage displays the combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd., and f1 are most often used in phage display libraries, as either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

For example, the recombinant phage antibody system (RPAS, Pharmacia Catalog number 27-9400-01) can be easily modified for use in expressing and screening *pan-s/tk* combinatorial libraries by panning on pancreatic β cells to enrich, in the flow through, for *pan-s/tk* homologs with enhanced ability to bind the ligand.

The invention also provides for reduction of the *pan-s/tk* protein to generate mimetics, e.g. peptide or non-peptide agents, which are able to disrupt a biological activity of a *pan-s/tk* polypeptide of the present invention, e.g. as inhibitors of protein-protein interactions, such as with ligand proteins. Thus, such mutagenic techniques as described above are also useful to map the determinants of the *pan-s/tk* proteins which participate in protein-protein interactions involved in, for example, interaction of the subject *pan-s/tk* polypeptide with ligand or alternatively with intracellular elements.

To illustrate, the critical residues of a subject *pan-s/tk* polypeptide which are involved in molecular recognition of a ligand can be determined and used to generate *pan-s/tk*-derived peptidomimetics which competitively inhibit binding of the authentic *pan-s/tk* protein with that moiety. By employing, for example, scanning mutagenesis to map the amino acid residues of a protein which is involved in binding other proteins, peptidomimetic compounds can be generated which mimic those residues which facilitate

the interaction. Such mimetics may then be used to interfere with the normal function of a *pan-s/tk* protein (or its ligand). For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in Peptides: Structure and Function (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), β-turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc* Perkin Trans 1:1231), and β-aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun* 126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

Another aspect of the invention pertains to an antibody specifically reactive with a *pan-s/tk* protein. For example, by using immunogens derived from a *pan-s/tk* protein, e.g. based on the cDNA sequences, anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, Antibodies: A Laboratory Manual ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal, such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide (e.g., a *pan-s/tk* polypeptide or an antigenic fragment which is capable of eliciting an antibody response). Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of a *pan-s/tk* protein can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies. In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of a *pan-s/tk* protein of a organism, such as a mammal, e.g. antigenic determinants of a protein represented by SEQ ID No. 2, 4, 6 or 9 or closely related homologs (e.g. at least 70% homologous, preferably at least 80% homologous, and more preferably at least 90% homologous). In yet a further preferred embodiment of the present invention, in order to provide, for example, antibodies which are immuno-selective for discrete *pan-s/tk* homologs the anti-*pan-s/tk* polypeptide antibodies do not substantially cross react (i.e. does not react specifically) with a protein which is, for example, less than 85%, 90% or 95% homologous with the selected *pan-s/tk*. By "not substantially cross react", it is meant that the antibody has a binding affinity for a non-homologous protein which is at least one order of magnitude, more preferably at least 2 orders of magnitude, and even more preferably at least 3 orders of magnitude less than the binding affinity of the antibody for the intended target *pan-s/tk*.

Following immunization of an animal with an antigenic preparation of a *pan-s/tk* polypeptide, anti-*pan-s/tk* antisera can be obtained and, if desired, polyclonal anti-*pan-s/tk* antibodies isolated from the serum. To produce monoclonal antibodies, antibody-producing cells (lymphocytes) can be harvested from an immunized animal and fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Such techniques are well known in the art, an include, for example, the hybridoma technique (originally developed by Kohler and Milstein, (1975) Nature, 256: 495-497), the human B cell hybridoma technique (Kozbar et al., (1983) Immunology Today, 4: 72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., (1985) Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. pp. 77-96). Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with a *pan-s/tk* polypeptide of the present invention and monoclonal antibodies isolated from a culture comprising such hybridoma cells.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with a *pan-s/tk* polypeptide. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab)_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab)_2$ fragment can be treated to reduce disulfide bridges to produce Fab fragments. The antibody of the present invention is further intended to include bispecific and chimeric molecules having affinity for a *pan-s/tk* protein conferred by at least one CDR region of the antibody.

Both monoclonal and polyclonal antibodies (Ab) directed against authentic *pan-s/tk* polypeptides, or *pan-s/tk* variants, and antibody fragments such as Fab, $F(ab)_2$, Fv and scFv can be used to block the action of a *pan-s/tk* protein and allow the study of the role of these proteins in, for example, differentiation of tissue. Experiments of this nature can aid in deciphering the role of *pan-s/tk* proteins that may be involved in control of proliferation versus differentiation, e.g., in patterning and tissue formation.

Antibodies which specifically bind *pan-s/tk* epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of each of the subject *pan-s/tk* polypeptides. Anti-*pan-s/tk* antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate *pan-s/tk* protein levels in tissue as part of a clinical testing procedure. For instance, such measurements can be useful in predictive valuations of the onset or progression of proliferative or differentiative disorders. Likewise, the ability to monitor *pan-s/tk* protein levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with such a disorder. The level of *pan-s/tk* polypeptides may be measured from cells in bodily fluid, such as in samples of cerebral spinal fluid or amniotic fluid, or can be measured in tissue, such as produced by biopsy. Diagnostic assays using

anti-*pan-s/tk* antibodies can include, for example, immunoassays designed to aid in early diagnosis of a disorder, particularly ones which are manifest at birth. Diagnostic assays using anti-*pan-s/tk* polypeptide antibodies can also include immunoassays designed to aid in early diagnosis and phenotyping neoplastic or hyperplastic disorders.

Another application of anti-*pan-s/tk* antibodies of the present invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λ gt11, λgt18-23, λZAP, and λORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λgt11 will produce fusion proteins whose amino termini consist of ß-galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a *pan-s/tk* protein, e.g. orthologs of the *pan-s/tk* protein from other species, can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-*pan-s/tk* antibodies. Positive phage detected by this assay can then be isolated from the infected plate. Thus, the presence of *pan-s/tk* homologs can be detected and cloned from other animals, as can alternate isoforms (including splicing variants) from humans.

Moreover, the nucleotide sequences determined from the cloning of *pan-s/tk* genes from organisms will further allow for the generation of probes and primers designed for use in identifying and/or cloning *pan-s/tk* homologs in other cell types, e.g. from other tissues, as well as *pan-s/tk* homologs from other organisms. For instance, the present invention also provides a probe/primer comprising a substantially purified oligonucleotide, which oligonucleotide comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least 15 consecutive nucleotides of sense or anti-sense sequence selected from the group consisting of SEQ ID No. 1, 3, 5, 7 or 8 or naturally occurring mutants thereof. For instance, primers based on the nucleic acid represented in SEQ ID No. 1, 3, 5, 7 or 8 can be used in PCR reactions to clone *pan-s/tk* homologs. Likewise, probes based on the subject *pan-s/tk* sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto and able to be detected, e.g. the label group is selected from amongst radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors.

Such probes can also be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a *pan-s/tk* protein, such as by measuring a level of a *pan-s/tk*-encoding nucleic acid in a sample of cells from a patient-animal; e.g. detecting *pan-s/tk* mRNA levels or determining whether a genomic *pan-s/tk* gene has been mutated or deleted.

To illustrate, nucleotide probes can be generated from the subject *pan-s/tk* genes which facilitate histological screening of intact tissue and tissue samples for the presence (or absence) of *pan-s/tk*-encoding transcripts. Similar to the diagnostic uses of anti-*pan-s/tk*

antibodies, the use of probes directed to *pan-s/tk* messages, or to genomic *pan-s/tk* sequences, can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, degenerative disorders marked by loss of particular cell-types, apoptosis, neoplastic and/or hyperplastic disorders (e.g. unwanted cell growth) or abnormal differentiation of tissue. Used in conjunction with immunoassays as described above, the oligonucleotide probes can help facilitate the determination of the molecular basis for a developmental disorder which may involve some abnormality associated with expression (or lack thereof) of a *pan-s/tk* protein. For instance, variation in polypeptide synthesis can be differentiated from a mutation in a coding sequence.

Accordingly, the present method provides a method for determining if a subject is at risk for a disorder characterized by aberrant apoptosis, cell proliferation and/or differentiation. In preferred embodiments, method can be generally characterized as comprising detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of (i) an alteration affecting the integrity of a gene encoding a *pan-s/tk*-protein, or (ii) the mis-expression of the *pan-s/tk* gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a *pan-s/tk* gene, (ii) an addition of one or more nucleotides to a *pan-s/tk* gene, (iii) a substitution of one or more nucleotides of a *pan-s/tk* gene, (iv) a gross chromosomal rearrangement of a *pan-s/tk* gene, (v) a gross alteration in the level of a messenger RNA transcript of a *pan-s/tk* gene, (vii) aberrant modification of a *pan-s/tk* gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a *pan-s/tk* gene, (viii) a non-wild type level of a *pan-s/tk*-protein, and (ix) inappropriate post-translational modification of a *pan-s/tk*-protein. As set out below, the present invention provides a large number of assay techniques for detecting lesions in a *pan-s/tk* gene, and importantly, provides the ability to discern between different molecular causes underlying *pan-s/tk*-dependent aberrant cell growth, proliferation and/or differentiation.

In an exemplary embodiment, there is provided a nucleic acid composition comprising a (purified) oligonucleotide probe including a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of a *pan-s/tk* gene, such as represented by SEQ ID No. 1, 3, 5, 7 or 8, or naturally occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with the subject *pan-s/tk* genes or naturally occurring mutants thereof. The nucleic acid of a cell is rendered accessible for hybridization, the probe is exposed to nucleic acid of the sample, and the hybridization of the probe to the sample nucleic acid is detected. Such techniques can be used to detect lesions at either the genomic or mRNA level, including deletions, substitutions, etc., as well as to determine mRNA transcript levels.

In certain embodiments, detection of the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) Science 241:1077-1080; and Nakazawa et al. (1944) PNAS 91:360-364), the later of which can be particularly useful for detecting point mutations in the *pan-s/tk* gene. In a merely illustrative embodiment, the method includes the steps of (i) collecting a sample of cells from a patient, (ii) isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, (iii) contacting the nucleic acid sample with one or more primers which specifically hybridize to a *pan-s/tk* gene under conditions such that hybridization and amplification of the *pan-s/tk* gene (if present) occurs, and (iv) detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample.

In still another embodiment, the level of a *pan-s/tk* protein can be detected by immunoassay. For instance, the cells of a biopsy sample can be lysed, and the level of a *pan-s/tk*-protein present in the cell can be quantitated by standard immunoassay techniques. In yet another exemplary embodiment, aberrant methylation patterns of a *pan-s/tk* gene can be detected by digesting genomic DNA from a patient sample with one or more restriction endonucleases that are sensitive to methylation and for which recognition sites exist in the *pan-s/tk* gene (including in the flanking and intronic sequences). See, for example, Buiting et al. (1994) Human Mol Genet 3:893-895. Digested DNA is separated by gel electrophoresis, and hybridized with probes derived from, for example, genomic or cDNA sequences. The methylation status of the *pan-s/tk* gene can be determined by comparison of the restriction pattern generated from the sample DNA with that for a standard of known methylation.

In still other embodiments, the extracellular domain of the *pan-s/tk* receptor can be used to quantitatively detect the level of *pan-s/tk* ligands. To illustrate, a soluble form of the N-terminus (extracellular domain) of the receptor can be generated by truncation of the protein prior to (N-terminal to) the transmembrane domain. Samples of bodily fluid(s), e.g., plasma, serum, lymph, marrow, cerebral/spinal fluid, urine and the like can be contacted with the receptor under conditions wherein ligand/receptor binding can occur, and the level of ligand/receptor complexes formed can be detected by any of a variety of techniques known in the art. For example, competitive binding assays using standardized samples of a known *pan-s/tk* ligand can be used to quantitate the amount of analyte bound from the fluid sample.

In yet other embodiments, such *pan-s/tk* receptors can be used to detect the presence of a *pan-s/tk* ligand on a cell surface. For instance, the *pan-s/tk* protein can be contacted with cells from a biopsy, and the ability of the *pan-s/tk* protein to decorate certain cells of the sample is ascertained. The binding of the *pan-s/tk* protein to cell populations of the

sample can be detected, for example, by the use of antibodies against the *pan-s/tk* protein, or by detection of a label associated with the *pan-s/tk* protein. In the case of the latter, the *pan-s/tk* protein can be labeled, for example, by chemical modification or as a fusion protein. Exemplary labels include radioisotopes, fluorescent compounds, enzyme co-factors, which can be added by chemical modification of the protein, and epitope tags such as myc, pFLAG and the like, or enzymatic activities such as GST or alkaline phosphatase which can be added either by chemical modification or by generation of a fusion protein.

Furthermore, the present invention also contemplates the detection of soluble forms of the *pan-s/tk* receptor in bodily fluid samples. As described in the art, e.g., see Diez-Ruiz et al. (1995) *Eur J Haematol* 54:1-8 and Owen-Schaub et al. (1995) *Cancer Lett* 94:1-8, in certain instances soluble forms of receptors are believed to play a role as modulators of the biological function of their cognate ligands in an agonist/antagonist pattern. In various pathologic states, the production and release of soluble *pan-s/tk* receptors may mediate host response and determine the course and outcome of disease by interacting with *pan-s/tk* ligands and competing with cell surface receptors. The determination of soluble *pan-s/tk* receptors in body fluids is a new tool to gain information about various disease states, and may be of prognostic value to a clinician. For example, the level of soluble *pan-s/tk* protein in a body fluid may give useful information for monitoring, *inter alia*, neurodegenerative disorders and/or pancreodegenerative diseases.

The level of soluble receptor present in a given sample can be quantitated, in light of the present disclosure, using known procedures and techniques. For example, antibodies immunoselective for the extracellular domain of the *pan-s/tk* protein can be used to detect and quantify its presence in a sample, e.g., by well-known immunoassay techniques. Alternatively, a labeled ligand of the receptor can be used to detect the presence of the receptor in the fluid sample.

In yet another aspect of the invention, the subject *pan-s/tk* polypeptides can be used to generate a "two hybrid" assay or an "interaction trap" assay (see, for example, U.S. Patent No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J Biol Chem 268:12046-12054; Bartel et al. (1993) Biotechniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; and Brent WO94/10300), for isolating coding sequences for other cellular proteins which bind *pan-s/tk*s ("*pan-s/tk*-binding proteins" or "*pan-s/tk*-bp"). Such *pan-s/tk*-binding proteins would likely be involved in the regulation of *pan-s/tk*, e.g., as sMAD proteins or other signal transducers.

Briefly, the interaction trap relies on reconstituting *in vivo* a functional transcriptional activator protein from two separate fusion proteins. In particular, the method makes use of chimeric genes which express hybrid proteins. To illustrate, a first hybrid gene comprises the coding sequence for a DNA-binding domain of a transcriptional activator fused in frame to the coding sequence for a *pan-s/tk* polypeptide, such as the

cytoplasmic domain. Preferably, if the kinase domain is included, one or more of the active site residues will be mutated to provide a catalytically inactive mutant which nevertheless retains the ability to bind to its intracellular substrate(s). The second hybrid protein encodes a transcriptional activation domain fused in frame to a sample gene from a cDNA library. If the bait and sample hybrid proteins are able to interact, e.g., form a *pan-s/tk*-dependent complex, they bring into close proximity the two domains of the transcriptional activator. This proximity is sufficient to cause transcription of a reporter gene which is operably linked to a transcriptional regulatory site responsive to the transcriptional activator, and expression of the reporter gene can be detected and used to score for the interaction of the *pan-s/tk* and sample proteins.

A number of techniques exist in the art for now identifying the ligand of the *pan-s/tk* receptor. For instance, expression cloning can be carried out on a cDNA or genomic library by isolating cells which are decorated with a labeled form of the receptor. In a preferred embodiment, the technique uses the *pan-s/tk* receptor in an *in situ* assay for detecting *pan-s/tk* ligands in tissue samples and whole organisms. In general, the RAP-*in situ* assay described below (for Receptor Affinity Probe) of Flanagan and Leder (see PCT publications WO 92/06220; and also Cheng et al. (1994) *Cell* 79:157-168) involves the use of an expression cloning system whereby a *pan-s/tk* ligand is scored on the basis of binding to a *pan-s/tk*/alkaline phosphatase fusion protein. In general, the method comprises (i) providing a hybrid molecule (the affinity probe) including the *pan-s/tk* receptor, or at least the extracellular domain thereof, covalently bonded to an enzymatically active tag, preferably for which chromogenic substrates exist, (ii) contacting the tissue or organism with the affinity probe to form complexes between the probe and a cognate ligand in the sample, removing unbound probe, and (iii) detecting the affinity complex using a chromogenic substrate for the enzymatic activity associated with the affinity probe.

This method, unlike other prior art methods which are carried out only on dispersed cell cultures, provides a means for probing non-dispersed and wholemount tissue and animal samples. The method can be used, in addition to facilitating the cloning of *pan-s/tk* ligands, also for detecting patterns of expression for particular ligands of the *pan-s/tk* receptor, for measuring the affinity of receptor/ligand interactions in tissue samples, as well as for generating drug screening assays in tissue samples. Moreover, the affinity probe can also be used in diagnostic screening to determine whether a *pan-s/tk* ligand is misexpressed.

Furthermore, by making available purified and recombinant *pan-s/tk* polypeptides, the present invention facilitates the development of assays which can be used to screen for drugs which are either agonists or antagonists of the normal cellular function of the subject *pan-s/tk* receptor, or of its role in the pathogenesis of cellular maintenance, differentiation and/or proliferation and disorders related thereto. In a general sense, the assay evaluates the ability of a compound to modulate binding between a *pan-s/tk* polypeptide and a molecule,

be it derived from a cellular protein (substrate or other intracellular signalling molecule) or an extracellular protein (ligand), that interacts with the *pan-s/tk* polypeptide. Exemplary compounds which can be screened against such *pan-s/tk*-mediated interactions include peptides, nucleic acids, carbohydrates, small organic molecules, and natural product extract libraries, such as isolated from animals, plants, fungus and/or microbes.

It is contemplated that any of the novel interactions described herein could be exploited in a drug screening assay. For example, in one embodiment, the interaction between a *pan-s/tk* protein and a ligand on the surface of a β cell can be detected in the presence and the absence of a test compound. Likewise, the ability of test compound to inhibit the kinase activity of the *pan-s/tk* polypeptide.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with upstream or downstream elements. Accordingly, in an exemplary screening assay of the present invention, a reaction mixture is generated to include a *pan-s/tk* polypeptide, compound(s) of interest, and a "target molecule", e.g., a protein, which interacts with the *pan-s/tk* polypeptide. Exemplary target molecules include ligands, as well as peptide and non-peptide substrates. Detection and quantification of interaction of the pan-s/tk polypeptide with the target molecule provides a means for determining a compound's efficacy at inhibiting (or potentiating) interaction between the *pan-s/tk* and the target molecule. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, interaction of the *pan-s/tk* polypeptide and target molecule is quantitated in the absence of the test compound.

Interaction between the *pan-s/tk* polypeptide and the target molecule may be detected by a variety of techniques. Modulation of the formation of complexes can be quantitated using, for example, detectably labeled proteins such as radiolabeled, fluorescently labeled, or enzymatically labeled *pan-s/tk* polypeptides, by immunoassay, by chromatographic detection, or by detecting the intrinsic activity of the kinase.

Typically, it will be desirable to immobilize either *pan-s/tk* or the target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins,

as well as to accommodate automation of the assay. Binding of *pan-s/tk* to the target molecule, in the presence and absence of a candidate agent, can be accomplished in any vessel suitable for containing the reactants. Examples include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/*pan-s/tk* (GST/*pan-s/tk*) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates, e.g. an ^{35}S-labeled, and the test compound, and the mixture incubated under conditions conducive to complex formation, e.g. at physiological conditions for salt and pH, though slightly more stringent conditions may be desired. Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly (e.g. beads placed in scintillant), or in the supernatant after the complexes are subsequently dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of target molecule found in the bead fraction quantitated from the gel using standard electrophoretic techniques such as described in the appended examples.

Other techniques for immobilizing proteins and other molecules on matrices are also available for use in the subject assay. For instance, either *pan-s/tk* or target molecule can be immobilized utilizing conjugation of biotin and streptavidin. For instance, biotinylated *pan-s/tk* molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with *pan-s/tk*, but which do not interfere with the interaction between the *pan-s/tk* and target molecule, can be derivatized to the wells of the plate, and *pan-s/tk* trapped in the wells by antibody conjugation. As above, preparations of an target molecule and a test compound are incubated in the *pan-s/tk*-presenting wells of the plate, and the amount of complex trapped in the well can be quantitated. Exemplary methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the target molecule, or which are reactive with *pan-s/tk* protein and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule, either intrinsic or extrinsic activity. In the instance of the latter, the enzyme can be chemically conjugated or provided as a fusion protein with the target molecule. To illustrate, the target molecule can be chemically cross-linked or genetically fused (if it is a polypeptide) with horseradish peroxidase, and the amount of polypeptide trapped in the complex can be assessed with a chromogenic substrate of the enzyme, e.g. 3,3'-diamino-benzadine terahydrochloride or 4-chloro-1-napthol. Likewise, a fusion protein comprising the polypeptide and glutathione-S-transferase can be

provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2,4-dinitrobenzene (Habig et al (1974) J Biol Chem 249:7130).

For processes which rely on immunodetection for quantitating proteins trapped in the complex, antibodies against the protein, such as anti-*pan-s/tk* antibodies, can be used. Alternatively, the protein to be detected in the complex can be "epitope tagged" in the form of a fusion protein which includes, in addition to the *pan-s/tk* sequence, a second polypeptide for which antibodies are readily available (e.g. from commercial sources). For instance, the GST fusion proteins described above can also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include myc-epitopes (e.g., see Ellison et al. (1991) J Biol Chem 266:21150-21157) which includes a 10-residue sequence from c-myc, as well as the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharamacia, NJ).

In another embodiment of a drug screening, a two hybrid assay (described *supra*) can be generated with a *pan-s/tk* and target molecule. Drug dependent inhibition or potentiation of the interaction can be scored.

In still other embodiments, the target molecule can be a substrate for the kinase activity of the pan-s/tk protein. The extent to which the substrate is converted to product in the presence of the test compound is compared with the extent of substrate conversion in the absence of the compound. This method is a simple and rapid screening test which, in one embodiment, uses a serine/threonine kinase pseudosubstrate peptide, the generation of which are well known in the art. In one embodiment, the phosphorylation of a substrate of *pan-s/tk* can be detected by radiolabeled phosphates, e.g., [^{32}P]-ATP. In other embodiments, the measurement of the kinase activity can be made by separation of the non-phosphorylated and phosphorylated forms of the peptide by use of high pressure liquid chromatography (HPLC). Still another means for detecting phosphorylation of peptide substrate is through detection by using anti-phosphoserine and anti-phosphothreonine antibodies. In yet another embodiment, the peptide susbtrate is modified by placing an (o-NO2)-tyrosine residue on the N-terminal side of the phosphorylated serine. This modification generally does not interfere with the ability of the peptide to be a kinase substrate, and causes phosphorylation to alter the absorbance of the peptide at 430 nm, which can be continually measured by spectrophotometric techniques.

Other spectrophotometric assays for kinase activity have been developed using coupled reactions. The conversion of phosphoenolpyruvate to pyruvate can occur in the presence of ADP generated by kinase phosphotransfer and pyruvate kinase. The pyruvate is then converted to lactate by lactate dehydrogenase and detected by reading the absorbance at 340 nm.

In yet another embodiment, the drug screening assay is derived to include a whole cell recombinantly expressing a *pan-s/tk* polypeptide. The ability of a test agent to alter the activity of the *pan-s/tk* protein can be detected by analysis of the recombinant cell. For example, agonists and antagonists of the *pan-s/tk* biological activity can by detected by scoring for alterations in growth or differentiation (phenotype) of the cell. General techniques for detecting each are well known, and will vary with respect to the source of the particular reagent cell utilized in any given assay.

In an exemplary embodiment, a cell which expresses the *pan-s/tk* receptor, e.g, whether endogenous or heterologous, can be contacted with a ligand of the *pan-s/tk* receptor which is capable of inducing signal transduction from the receptor, and the resulting signaling detected either at various points in the pathway, or on the basis of a phenotypic change to the reagent cell. In one embodiment, the reagent cell is contacted with antibody which causes cross-linking of the receptor, and the signal cascade induced by that cross-linking is subsequently detected. A test compound which modulates that pathway, e.g., potentiates or inhibits, can be detected by comparison with control experiments which either lack the receptor or lack the test compound. For example, visual inspection of the morphology of the reagent cell can be used to determine whether the biological activity of the targeted *pan-s/tk* protein has been affected by the added agent. In yet another embodiment, the assay can be generated to evolve a detection signal from the expression or modification of a cellular protein effected by the activity of *pan-s/tk*-mediated signaling. Such measurement can be accomplished by detecting a biological activity modulated by the downstream effects of the receptor activity.

For example, the alteration of expression of a reporter gene construct provided in the reagent cell provides a means of detecting the effect on *pan-s/tk* activity. For example, reporter gene constructs derived using the transcriptional regulatory sequences, e.g. the promoters, from genes regulated by the signalling of the *pan-s/tk receptor* can be used to drive the expression of a detectable marker. Many reporter genes are known to those of skill in the art and others may be identified or synthesized by methods known to those of skill in the art. A reporter gene includes any gene that expresses a detectable gene product, which may be RNA or protein. Preferred reporter genes are those that are readily detectable. The reporter gene may also be included in the construct in the form of a fusion gene with a gene that includes desired transcriptional regulatory sequences or exhibits other desirable properties. Examples of reporter genes include, but are not limited to CAT (chloramphenicol acetyl transferase) (Alton and Vapnek (1979), Nature 282: 864-869) luciferase, and other enzyme detection systems, such as beta-galactosidase; firefly luciferase (deWet et al. (1987), Mol. Cell. Biol. 7:725-737); bacterial luciferase (Engebrecht and Silverman (1984), PNAS 1: 4154-4158; Baldwin et al. (1984), Biochemistry 23: 3663-3667); alkaline phosphatase (Toh et al. (1989) Eur. J. Biochem. 182: 231-238, Hall et al.

(1983) J. Mol. Appl. Gen. 2: 101), human placental secreted alkaline phosphatase (Cullen and Malim (1992) Methods in Enzymol. 216:362-368).

In still other embodiments, the signal generated by engagement of the *pan-s/tk* receptor can be detected by scoring for the production of second messengers. For example, in various embodiments the assay may assess the ability of test agent to cause changes in phophorylation patterns, adenylate cyclase activity (cAMP production), GTP hydrolysis, calcium mobilization, and/or phospholipid hydrolysis upon receptor stimulation.

Another aspect of the present invention relates to a method of inducing and/or maintaining a differentiated state, enhancing survival, and/or inhibiting (or alternatively potentiating) proliferation of a cell, by contacting the cells with an agent which modulates *pan-s/tk*-dependent signal transduction pathways. The subject method could be used to generate and/or maintain an array of different tissue both *in vitro* and *in vivo*. A "*pan-s/tk* therapeutic", whether inhibitory or potentiating with respect to modulating signaling by the *pan-s/tk* receptor, can be, as appropriate, any of the preparations described above, including isolated polypeptides, gene therapy constructs, antisense molecules, peptidomimetics or agents identified in the drug assays provided herein. In certain embodiments, soluble forms of the *pan-s/tk* protein including the extracellular ligand-binding domain of the receptor can be provided as a means for antagonizing the binding of a *pan-s/tk* ligand to a cell-surface *pan-s/tk* receptor. For instance, such forms of the receptor can be used to antagonize the bioactivity of a ligand of the receptor. In other embodiments, the *pan-s/tk* therapeutic can be an expression vector encoding a constitutively active kinase domain of the subject receptor.

The *pan-s/tk* compounds of the present invention are likely to play an important role in the modulation of cellular proliferation and maintenance of, e.g., pancreatic, neuronal, kidney and heart tissues during developmental and disease states. It will also be apparent that, by transient use of modulators of *pan-s/tk* activities, *in vivo* reformation of tissue can be accomplished, e.g. in the development and maintenance of organs. By controlling the proliferative and differentiative potential for different cells, the subject *pan-s/tk* therapeutics can be used to reform injured tissue, or to improve grafting and morphology of transplanted tissue. For instance, *pan-s/tk* antagonists and agonists can be employed in a differential manner to regulate different stages of organ repair after physical, chemical or pathological insult. The present method is also applicable to cell culture techniques.

In one embodiment, a *pan-s/tk* therapeutic of the present invention can be used to induce differentiation of uncommitted pancreatic or neuronal progenitor cells and thereby give rise to a committed progenitor cell, or to cause further restriction of the developmental fate of a committed progenitor cell towards becoming a particular terminally-differentiated cell.

Another aspect of the invention features transgenic non-human animals which express a heterologous *pan-s/tk* gene of the present invention, and/or which have had one or more genomic *pan-s/tk* genes disrupted in at least a tissue or cell-types of the animal. Accordingly, the invention features an animal model for developmental diseases, which animal has one or more *pan-s/tk* allele which is mis-expressed. For example, an animal can be generated which has one or more *pan-s/tk* alleles deleted or otherwise rendered inactive. Such a model can then be used to study disorders arising from mis-expressed *pan-s/tk* genes, as well as for evaluating potential therapies for similar disorders.

The transgenic animals of the present invention all include within a plurality of their cells a transgene of the present invention, which transgene alters the phenotype of the "host cell" with respect to regulation by the *pan-s/tk* receptor, e.g., of cell growth, death and/or differentiation. Since it is possible to produce transgenic organisms of the invention utilizing one or more of the transgene constructs described herein, a general description will be given of the production of transgenic organisms by referring generally to exogenous genetic material. This general description can be adapted by those skilled in the art in order to incorporate specific transgene sequences into organisms utilizing the methods and materials described below.

In one embodiment, the transgene construct is a knockout construct. Such transgene constructs usually are insertion-type or replacement-type constructs (Hasty et al. (1991) *Mol Cell Biol* 11:4509). The transgene constructs for disruption of a *pan-s/tk* gene are designed to facilitate homologous recombination with a portion of the genomic *pan-s/tk* gene so as to prevent the functional expression of the endogenous *pan-s/tk* gene. In preferred embodiments, the nucleotide sequence used as the knockout construct can be comprised of (1) DNA from some portion of the endogenous *pan-s/tk* gene (exon sequence, intron sequence, promoter sequences, etc.) which direct recombination and (2) a marker sequence which is used to detect the presence of the knockout construct in the cell. The knockout construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to prevent or interrupt transcription of the native *pan-s/tk* gene. Such insertion can occur by homologous recombination, i.e., regions of the knockout construct that are homologous to the endogenous *pan-s/tk* gene sequence hybridize to the genomic DNA and recombine with the genomic sequences so that the construct is incorporated into the corresponding position of the genomic DNA. The knockout construct can comprise (1) a full or partial sequence of one or more exons and/or introns of the *pan-s/tk* gene to be disrupted, (2) sequences which flank the 5' and 3' ends of the coding sequence of the *pan-s/tk* gene, or (3) a combination thereof.

A preferred knockout construct will delete, by targeted homologous recombination, essential structural elements of an endogenous *pan-s/tk* gene. For example, the targeting

construct can recombine with the genomic *pan-s/tk* gene can delete a portion of the coding sequence, and/or essential transcriptional regulatory sequences of the gene.

Alternatively, the knockout construct can be used to interrupt essential structural and/or regulatory elements of an endogenous *pan-s/tk* gene by targeted insertion of a polynucleotide sequence. For instance, a knockout construct can recombine with a *pan-s/tk* gene and insert a nonhomologous sequence, such as a *neo* expression cassette, into a structural element (e.g., an exon) and/or regulatory element (e.g., enhancer, promoter, intron splice site, polyadenylation site, etc.) to yield a targeted *pan-s/tk* allele having an insertional disruption. The inserted nucleic acid can range in size from 1 nucleotide (e.g., to produce a frameshift) to several kilobases or more, and is limited only by the efficiency of the targeting technique.

Depending of the location and characteristics of the disruption, the transgene construct can be used to generate a transgenic animal in which substantially all expression of the targeted *pan-s/tk* gene is inhibited in at least a portion of the animal's cells. If only regulatory elements are targeted, some low-level expression of the targeted gene may occur (i.e., the targeted allele is "leaky").

The nucleotide sequence(s) comprising the knockout construct(s) can be obtained using methods well known in the art. Such methods include, for example, screening genomic libraries with *pan-s/tk* cDNA probes in order to identify the corresponding genomic *pan-s/tk* gene and regulatory sequences. Alternatively, where the cDNA sequence is to be used as part of the knockout construct, the cDNA may be obtained by screening a cDNA library as set out above.

In another embodiment, the "transgenic non-human animals" of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonal target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonal target cell. The specific line(s) of any animal used to practice this invention are selected for general good health, good embryo yields, good pronuclear visibility in the embryo, and good reproductive fitness. In addition, the haplotype is a significant factor. For example, when transgenic mice are to be produced, strains such as C57BL/6 or FVB lines are often used (Jackson Laboratory, Bar Harbor, ME). Preferred strains are those with H-2^b, H-2^d or H-2^q haplotypes such as C57BL/6 or DBA/1. The line(s) used to practice this invention may themselves be transgenics, and/or may be knockouts (i.e., obtained from animals which have one or more genes partially or completely suppressed) .

In one embodiment, the transgene construct is introduced into a single stage embryo. The zygote is the best target for micro-injection. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated

into the host gene before the first cleavage (Brinster et al. (1985) *PNAS* 82:4438-4442). As a consequence, all cells of the transgenic animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene.

Introduction of the transgene nucleotide sequence into the embryo may be accomplished by any means known in the art such as, for example, microinjection, electroporation, or lipofection. Following introduction of the transgene nucleotide sequence into the embryo, the embryo may be incubated *in vitro* for varying amounts of time, or reimplanted into the surrogate host, or both. In vitro incubation to maturity is within the scope of this invention. One common method in to incubate the embryos *in vitro* for about 1-7 days, depending on the species, and then reimplant them into the surrogate host.

Any technique which allows for the addition of the exogenous genetic material into nucleic genetic material can be utilized so long as it is not destructive to the cell, nuclear membrane or other existing cellular or genetic structures. The exogenous genetic material is preferentially inserted into the nucleic genetic material by microinjection. Microinjection of cells and cellular structures is known and is used in the art.

Reimplantation is accomplished using standard methods. Usually, the surrogate host is anesthetized, and the embryos are inserted into the oviduct. The number of embryos implanted into a particular host will vary by species, but will usually be comparable to the number of off spring the species naturally produces.

Transgenic offspring of the surrogate host may be screened for the presence and/or expression of the transgene by any suitable method. Screening is often accomplished by Southern blot or Northern blot analysis, using a probe that is complementary to at least a portion of the transgene. Western blot analysis using an antibody against the protein encoded by the transgene may be employed as an alternative or additional method for screening for the presence of the transgene product. Typically, DNA is prepared from excised tissue and analyzed by Southern analysis or PCR for the transgene. Alternatively, the tissues or cells believed to express the transgene at the highest levels are tested for the presence and expression of the transgene using Southern analysis or PCR, although any tissues or cell types may be used for this analysis.

Retroviral infection can also be used to introduce transgene into a non-human animal. The developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenich, R. (1976) *PNAS* 73:1260-1264). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (*Manipulating the Mouse Embryo*, Hogan eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the

transgene (Jahner et al. (1985) *PNAS* 82:6927-6931; Van der Putten et al. (1985) *PNAS* 82:6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, *supra*; Stewart et al. (1987) *EMBO J.* 6:383-388). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner et al. (1982) *Nature* 298:623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic non-human animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line by intrauterine retroviral infection of the midgestation embryo (Jahner et al. (1982) *supra*).

A third type of target cell for transgene introduction is the embryonal stem cell (ES). ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (Evans et al. (1981) *Nature* 292:154-156; Bradley et al. (1984) *Nature* 309:255-258; Gossler et al. (1986) *PNAS* 83: 9065-9069; and Robertson et al. (1986) *Nature* 322:445-448). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retrovirus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. For review see Jaenisch, R. (1988) *Science* 240:1468-1474.

In one embodiment, gene targeting, which is a method of using homologous recombination to modify an animal's genome, can be used to introduce changes into cultured embryonic stem cells. By targeting the *pan-s/tk* gene in ES cells, these changes can be introduced into the germlines of animals to generate chimeras. The gene targeting procedure is accomplished by introducing into tissue culture cells a DNA targeting construct that includes a segment homologous to a *pan-s/tk* locus, and which also includes an intended sequence modification to the *pan-s/tk* genomic sequence (e.g., insertion, deletion, point mutation). The treated cells are then screened for accurate targeting to identify and isolate those which have been properly targeted.

Gene targeting in embryonic stem cells is in fact a scheme contemplated by the present invention as a means for disrupting a *pan-s/tk* gene function through the use of a targeting transgene construct designed to undergo homologous recombination with *pan-s/tk* genomic sequences. Targeting construct can be arranged so that, upon recombination with an element of a *pan-s/tk* gene, a positive selection marker is inserted into (or replaces) coding sequences of the targeted *pan-s/tk* gene. The inserted sequence functionally disrupts the *pan-s/tk* gene, while also providing a positive selection trait.

Generally, the embryonic stem cells (ES cells) used to produce the knockout animals will be of the same species as the knockout animal to be generated. Thus for

example, mouse embryonic stem cells will usually be used for generation of a *pan-s/tk-* knockout mice.

Embryonic stem cells are generated and maintained using methods well known to the skilled artisan such as those described by Doetschman et al. (1985) *J. Embryol. Exp. Morphol.* 87:27-45). Any line of ES cells can be used, however, the line chosen is typically selected for the ability of the cells to integrate into and become part of the germ line of a developing embryo so as to create germ line transmission of the knockout construct. Thus, any ES cell line that is believed to have this capability is suitable for use herein. The cells are cultured and prepared for knockout construct insertion using methods well known to the skilled artisan, such as those set forth by Robertson in: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, E.J. Robertson, ed. IRL Press, Washington, D.C. [1987]); by Bradley et al. (1986) *Current Topics in Devel. Biol.* 20:357-371); and by Hogan et al. (Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY [1986]) .

Insertion of the knockout construct into the ES cells can be accomplished using a variety of methods well known in the art including for example, electroporation, microinjection, and calcium phosphate treatment. A preferred method of insertion is electroporation .

Each knockout construct to be inserted into the cell must first be in the linear form. Therefore, if the knockout construct has been inserted into a vector, linearization is accomplished by digesting the DNA with a suitable restriction endonuclease selected to cut only within the vector sequence and not within the knockout construct sequence.

For insertion, the knockout construct is added to the ES cells under appropriate conditions for the insertion method chosen, as is known to the skilled artisan. Where more than one construct is to be introduced into the ES cell, each knockout construct can be introduced simultaneously or one at a time.

If the ES cells are to be electroporated, the ES cells and knockout construct DNA are exposed to an electric pulse using an electroporation machine and following the manufacturer's guidelines for use. After electroporation, the ES cells are typically allowed to recover under suitable incubation conditions. The cells are then screened for the presence of the knockout construct .

Screening can be accomplished using a variety of methods. Where the marker gene is an antibiotic resistance gene, the ES cells may be cultured in the presence of an otherwise lethal concentration of antibiotic. Those ES cells that survive have presumably integrated the knockout construct. If the marker gene is other than an antibiotic resistance gene, a Southern blot of the ES cell genomic DNA can be probed with a sequence of DNA designed to hybridize only to the marker sequence Alternatively, PCR can be used. Finally, if the

marker gene is a gene that encodes an enzyme whose activity can be detected (e.g., β-galactosidase), the enzyme substrate can be added to the cells under suitable conditions, and the enzymatic activity can be analyzed. One skilled in the art will be familiar with other useful markers and the means for detecting their presence in a given cell. All such markers are contemplated as being included within the scope of the teaching of this invention.

The knockout construct may integrate into several locations in the ES cell genome, and may integrate into a different location in each ES cell's genome due to the occurrence of random insertion events. The desired location of insertion is in a complementary position to the DNA sequence to be knocked out, e.g., the *pan-s/tk* coding sequence, transcriptional regulatory sequence, etc. Typically, less than about 1-5 percent of the ES cells that take up the knockout construct will actually integrate the knockout construct in the desired location. To identify those ES cells with proper integration of the knockout construct, total DNA can be extracted from the ES cells using standard methods. The DNA can then be probed on a Southern blot with a probe or probes designed to hybridize in a specific pattern to genomic DNA digested with particular restriction enzyme(s). Alternatively, or additionally, the genomic DNA can be amplified by PCR with probes specifically designed to amplify DNA fragments of a particular size and sequence (i.e., only those cells containing the knockout construct in the proper position will generate DNA fragments of the proper size).

After suitable ES cells containing the knockout construct in the proper location have been identified, the cells can be inserted into an embryo. Insertion may be accomplished in a variety of ways known to the skilled artisan, however a preferred method is by microinjection. For microinjection, about 10-30 cells are collected into a micropipet and injected into embryos that are at the proper stage of development to permit integration of the foreign ES cell containing the knockout construct into the developing embryo. For instance, the transformed ES cells can be microinjected into blastocytes.

After the ES cell has been introduced into the embryo, the embryo may be implanted into the uterus of a pseudopregnant foster mother for gestation. While any foster mother may be used, the foster mother is typically selected for her ability to breed and reproduce well, and for her ability to care for the young. Such foster mothers are typically prepared by mating with vasectomized males of the same species. The stage of the pseudopregnant foster mother is important for successful implantation, and it is species dependent.

Offspring that are born to the foster mother may be screened initially for *pan-s/tk* disruptants, DNA from tissue of the offspring may be screened for the presence of the knockout construct using Southern blots and/or PCR as described above. Offspring that appear to be mosaics may then be crossed to each other, if they are believed to carry the knockout construct in their germ line, in order to generate homozygous knockout animals. Homozygotes may be identified by Southern blotting of equivalent amounts of genomic

DNA from animals that are the product of this cross, as well as animals that are known heterozygotes and wild type animals.

Other means of identifying and characterizing the knockout offspring are available. For example, Northern blots can be used to probe the mRNA for the presence or absence of transcripts of either the *pan-s/tk* gene, the marker gene, or both. In addition, Western blots can be used to assess the (loss of) level of expression of the *pan-s/tk* gene knocked out in various tissues of the offspring by probing the Western blot with an antibody against the *pan-s/tk* protein, or an antibody against the marker gene product, where this gene is expressed. Finally, *in situ* analysis (such as fixing the cells and labeling with antibody) and/or FACS (fluorescence activated cell sorting) analysis of various cells from the offspring can be conducted using suitable antibodies or *pan-s/tk* ligands to look for the presence or absence of the knockout construct gene product.

Animals containing more than one knockout construct and/or more than one transgene expression construct are prepared in any of several ways. The preferred manner of preparation is to generate a series of animals, each containing a desired transgenic phenotypes. Such animals are bred together through a series of crosses, backcrosses and selections, to ultimately generate a single animal containing all desired knockout constructs and/or expression constructs, where the animal is otherwise congenic (genetically identical) to the wild type except for the presence of the knockout construct(s) and/or transgene(s).

The transformed animals, their progeny, and cell lines of the present invention provide several important uses that will be readily apparent to one of ordinary skill in the art.

To illustrate, the transgenic animals and cell lines are particularly useful in screening compounds that have potential as prophylactic or therapeutic treatments of diseases such as may involve aberrant expression, or loss, of a *pan-s/tk* gene, or aberrant or unwanted activation of receptor signaling. Screening for a useful drug would involve administering the candidate drug over a range of doses to the transgenic animal, and assaying at various time points for the effect(s) of the drug on the disease or disorder being evaluated. Alternatively, or additionally, the drug could be administered prior to or simultaneously with exposure to induction of the disease, if applicable.

In one embodiment, candidate compounds are screened by being administered to the transgenic animal, over a range of doses, and evaluating the animal's physiological response to the compound(s) over time. Administration may be oral, or by suitable injection, depending on the chemical nature of the compound being evaluated. In some cases, it may be appropriate to administer the compound in conjunction with co-factors that would enhance the efficacy of the compound.

In screening cell lines derived from the subject transgenic animals for compounds useful in treating various disorders, the test compound is added to the cell culture medium at the appropriate time, and the cellular response to the compound is evaluated over time using the appropriate biochemical and/or histological assays. In some cases, it may be appropriate to apply the compound of interest to the culture medium in conjunction with co-factors that would enhance the efficacy of the compound.

All of the above-cited references and publications are hereby incorporated by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific polypeptides, nucleic acids, methods, assays and reagents described herein. Such equivalents are considered to be within the scope of this invention.

1. An isolated and/or recombinant *pan-s/tk* polypeptide comprising a *pan-s/tk* amino acid sequence identical or homologous to an amino acid sequence represented in SEQ ID No. 2, 4, 6 or 9.

2. An isolated and/or recombinant *pan-s/tk* polypeptide comprising a *pan-s/tk* amino acid sequence at least 70 percent identical to SEQ ID No. 2, 4, 6 or 9, or a portion thereof which retains a kinase activity of the *pan-s/tk* of SEQ ID No. 2, 4, 6 or 9.

3. An isolated and/or recombinant *pan-s/tk* polypeptide comprising an amino acid sequence encoded by a nucleic acid which hybridizes under stringent conditions to a mammalian *pan-s/tk* gene.

4. An isolated and/or recombinant *pan-s/tk* polypeptide comprising an amino acid sequence cross-reactive with an antibody specific for the *pan-s/tk* protein designated in SEQ ID No. 2, 4, 6 or 9.

5. The *pan-s/tk* polypeptide of any of claims 1, 2, 3 or 4, which polypeptide modulates at least one of proliferation, differentiation or survival of a cell which expresses the *pan-s/tk* polypeptide.

6. The *pan-s/tk* polypeptide of claim 5, wherein the cell is a pancreatic cell.

7. The *pan-s/tk* polypeptide of any of claims 1, 2, 3 or 4, which polypeptide comprises an amino acid sequence at least 75% homologous with the amino acid sequence designated by SEQ ID No. 2, 4, 6 or 9.

8. The *pan-s/tk* polypeptide of claim 7, which polypeptide comprises an amino acid sequence at least 85% homologous with the amino acid sequence designated by SEQ ID No. 2, 4, 6 or 9.

9. The *pan-s/tk* polypeptide of claim 7, which polypeptide comprises an amino acid sequence at least 95% homologous with the amino acid sequence designated by SEQ ID No. 2, 4, 6 or 9.

10. The *pan-s/tk* polypeptide of claim 7, which polypeptide comprises an amino acid sequence identical with the amino acid sequence designated by SEQ ID No. 2, 4, 6 or 9.

11. The *pan-s/tk* polypeptide of any of claims 1, 2, 3 or 4, which polypeptide comprises a serine/threonine kinase domain.

12. The *pan-s/tk* polypeptide of any of claims 1, 2, 3 or 4, which polypeptide is of mammalian origin.

13. The *pan-s/tk* polypeptide of any of claims 1, 2, 3 or 4, which polypeptide is a soluble polypeptide.

14. The *pan-s/tk* polypeptide of any of claims 1, 2, 3 or 4, which polypeptide is substantially free of other cellular proteins with each it naturally associates.

15. The *pan-s/tk* polypeptide of any of claims 1, 2, 3 or 4, which polypeptide is a fusion protein.

16. The *pan-s/tk* polypeptide of claim 15, wherein the fusion protein includes, as a second polypeptide sequence, a polypeptide which functions as a detectable label for detecting the presence of the fusion protein or as a matrix-binding domain for immobilizing the fusion protein.

17. An immunogen comprising the *pan-s/tk* polypeptide of claim 1, in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for the *pan-s/tk* polypeptide.

18. An antibody preparation specifically reactive with an epitope of the *pan-s/tk* polypeptide of claim 1.

19. An isolated nucleic acid comprising a coding sequence encoding a recombinant polypeptide comprising a *pan-s/tk* polypeptide sequence identical or homologous to an amino acid sequence represented in SEQ ID No. 2, 4, 6 or 9.

20. An isolated nucleic acid encoding a recombinant polypeptide comprising a *pan-s/tk* coding sequence which hybridizes to a mammalian *pan-s/tk* gene.

21. The nucleic acid of any of claims 19 or 20, which coding sequence hybridizes under stringent conditions to a nucleic acid probe having a sequence represented by at least 12 consecutive nucleotides of SEQ ID No. 1, 3, 5, 7 or 8.

22. The nucleic acid of any of claims 19 or 20, further comprising a transcriptional regulatory sequence operably linked to the coding sequence so as to render the nucleic acid suitable for use as an expression vector.

23. An expression vector, capable of replicating in at least one of a prokaryotic cell and eukaryotic cell, comprising the nucleic acid of claim 22.

24. A host cell transfected with the expression vector of claim 23 and expressing the recombinant polypeptide.

25. A method of producing a recombinant *pan-s/tk* polypeptide comprising culturing the cell of claim 24in a cell culture medium to cause expression of a *pan-s/tk* polypeptide encoded by the expression vector, and isolating the *pan-s/tk* polypeptide from the cell culture.

26. A transgenic animal having cells which harbor a transgene comprising the nucleic acid of claim 19.

27. A transgenic animal in which *pan-s/tk* stimulated signal transduction pathways are inhibited in one or more tissue of the animal by one of either expression of an antagonistic *pan-s/tk* polypeptide or disruption of a *pan-s/tk* gene.

28. A recombinant gene comprising a *pan-s/tk* encoding nucleotide sequence identical or homologous with SEQ ID No. 1, 3, 5, 7 or 8, or a fragment thereof, the nucleotide sequence operably linked to a transcriptional regulatory sequence in an open reading frame and translatable to a polypeptide.

29. The recombinant gene of claim 28, wherein the *pan-s/tk* encoding nucleotide sequence is derived from a genomic clone and includes intronic nucleotide sequences disrupting the open reading frame.

30. A nucleic acid comprising a substantially purified oligonucleotide, the oligonucleotide containing a region of nucleotide sequence which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence of SEQ ID No. 1, 3, 5, 7 or 8, or naturally occurring mutants thereof.

31. The nucleic acid of claim 30, which nucleic acid further comprises a label group attached thereto and able to be detected.

32. A test kit for detecting cells which contain a *pan-s/tk* mRNA transcript, comprising a nucleic acid of claim 30 for measuring, in a sample of cells, a level of nucleic acid encoding a *pan-s/tk* protein.

33. A test kit for detecting cells or tissue containing a *pan-s/tk* protein, comprising an antibody specific for a *pan-s/tk protein* for measuring, in a sample of cells, a level of the *pan-s/tk* protein.

34. A method for modulating, in an animal, cell growth, differentiation or survival, comprising administering a therapeutically effective amount of a *pan-s/tk* polypeptide.

35. The method of claim 34, comprising administering a nucleic acid construct encoding a *pan-s/tk* polypeptide under conditions wherein the construct is incorporated and recombinantly expressed by the cells to be modulated or cells located proximate thereto.

36. A recombinant transfection system, comprising
(i) a gene construct encoding a *pan-s/tk* polypeptide and operably linked to a transcriptional regulatory sequence for causing expression of the *pan-s/tk* polypeptide in eukaryotic cells, and
(ii) a gene delivery composition for delivering the gene construct to a cell and causing the cell to be transfected with the gene construct.

37. The recombinant transfection system of claim 36, wherein the gene delivery composition is selected from a group consisting of a recombinant viral particle, a liposome, and a poly-cationic nucleic acid binding agent,

38. A method of determining if a subject is at risk for a disorder characterized by unwanted cell proliferation, differentiation or death, comprising detecting, in a tissue of the subject, the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a *pan-s/tk* protein; and (ii) the mis-expression of the gene.

39. The method of claim 38, wherein detecting the genetic lesion comprises ascertaining the existence of at least one of
i. a deletion of one or more nucleotides from the gene,
ii. an addition of one or more nucleotides to the gene,
iii. an substitution of one or more nucleotides of the gene,
iv. a gross chromosomal rearrangement of the gene,

v. aberrant methylation of the gene,
vi. a gross alteration in the level of a messenger RNA transcript of the gene,
vii. the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene, and
viii. a non-wild type level of the protein.

40. The method of claim 38, wherein detecting the genetic lesion comprises
i. providing a nucleic acid comprising an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of SEQ ID No. 1, 3, 5, 7 or 8 or naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the gene;
ii. exposing the nucleic acid to nucleic acid of the tissue; and
iii. detecting, by hybridization of the nucleic acid to the nucleic acid, the presence or absence of the genetic lesion.

41. The method of claim 39, wherein detection of the genetic lesion comprises detecting the presence or absence of a *pan-s/tk* protein in cells of a tissue sample and/or as soluble proteins in bodily fluid.

42. A method of detecting the presence of a *pan-s/tk* ligand on cells present in a biological sample, comprising contacting the cells with a labeled *pan-s/tk* polypeptide and under conditions where the *pan-s/tk* polypeptide can specifically bind to cognate ligand, and detecting presence of the *pan-s/tk* polypeptide bound to the cells.

43. An assay for screening test compounds that modulate the bioactivity of a *pan-s/tk* receptor comprising:
i. combining a test compound, a *pan-s/tk* polypeptide, and a target compound selected from the group consisting of a *pan-s/tk* ligand, a signal transduction protein which binds to the *pan-s/tk* polypeptide, or a substrate of a kinase activity of the *pan-s/tk* polypeptide; and
ii. detecting the interaction of the target compound and the *pan-s/tk* polypeptide, wherein a change in the interaction of the target compound and the *pan-s/tk* polypeptide in the presence of the test compound is indicative of a potential ability to modulate the bioactivity of the *pan-s/tk* receptor.

44. The assay of claim 43, wherein the *pan-s/tk* polypeptide is a soluble polypeptide.

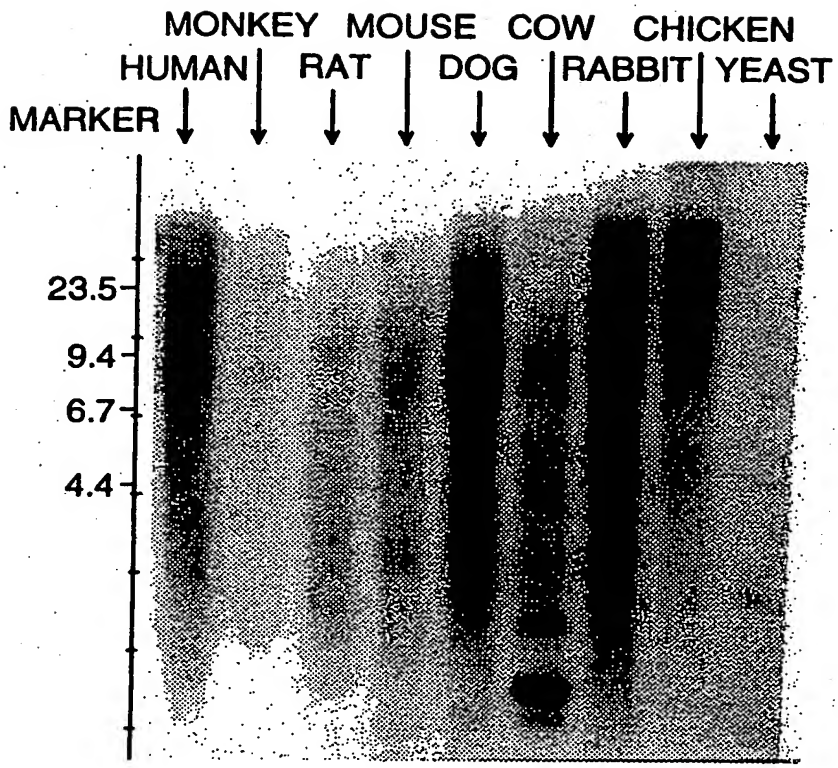
MONKEY
MOUSE
COW
CHICKEN
HUMAN
RAT
DOG
RABBIT
YEAST
MARKER
23.5
9.4
6.7
4.4

FIG. 1

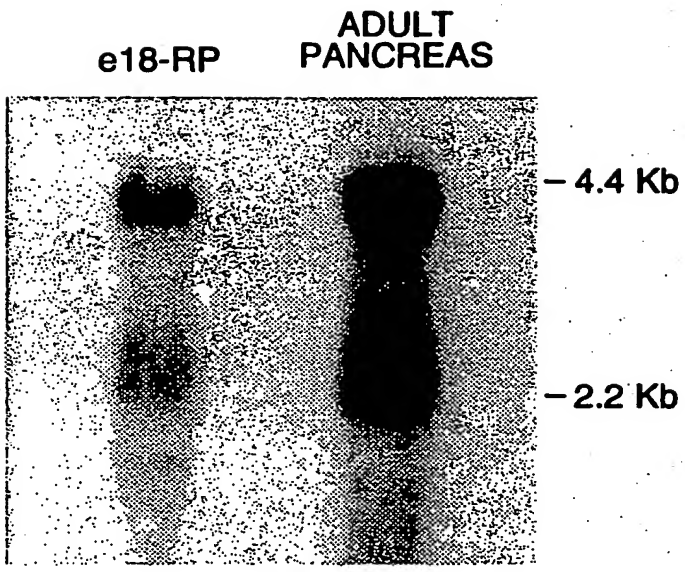
e18-RP
ADULT PANCREAS
4.4 Kb
2.2 Kb

FIG. 2A

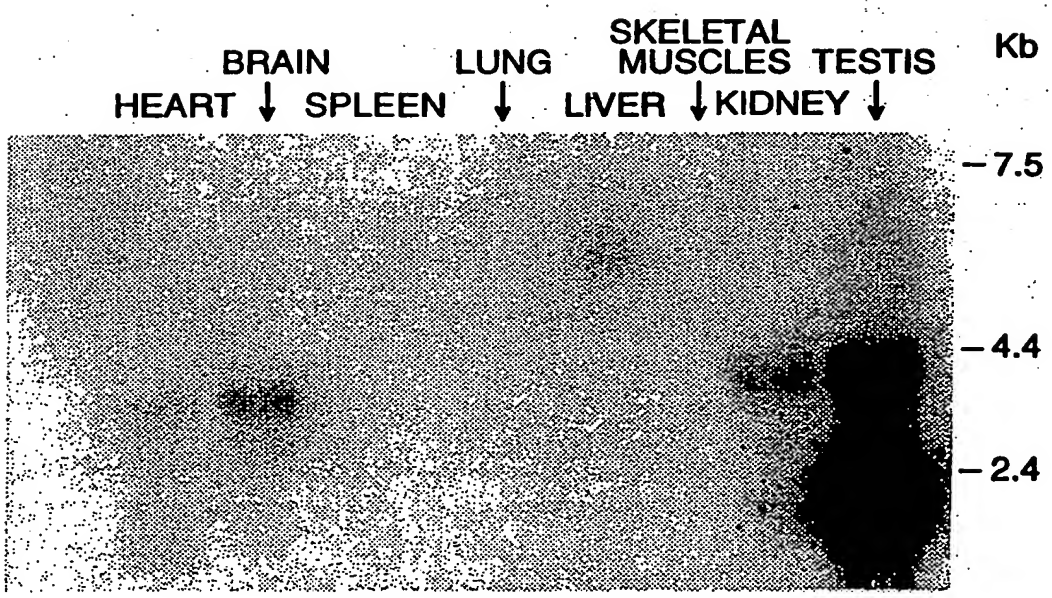
BRAIN
HEART
SPLEEN
LUNG
SKELETAL MUSCLES
LIVER
TESTIS
KIDNEY
Kb
7.5
4.4
2.4

FIG. 2B

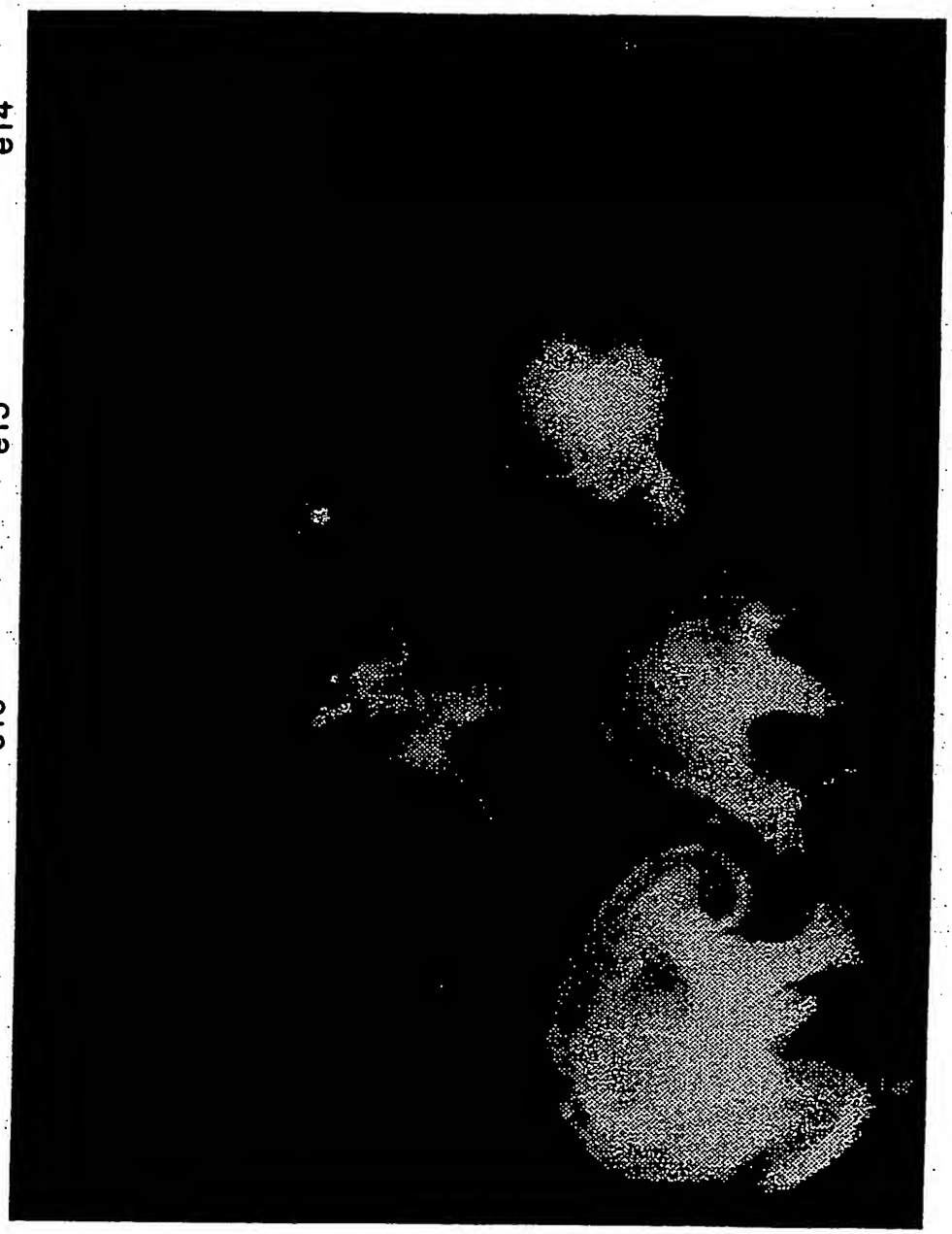
e14
e15
e16
e17
ANTISENSE (TOP)
SENSE CONTROL (BOTTOM)
WHOLE MOUNT

FIG. 3A

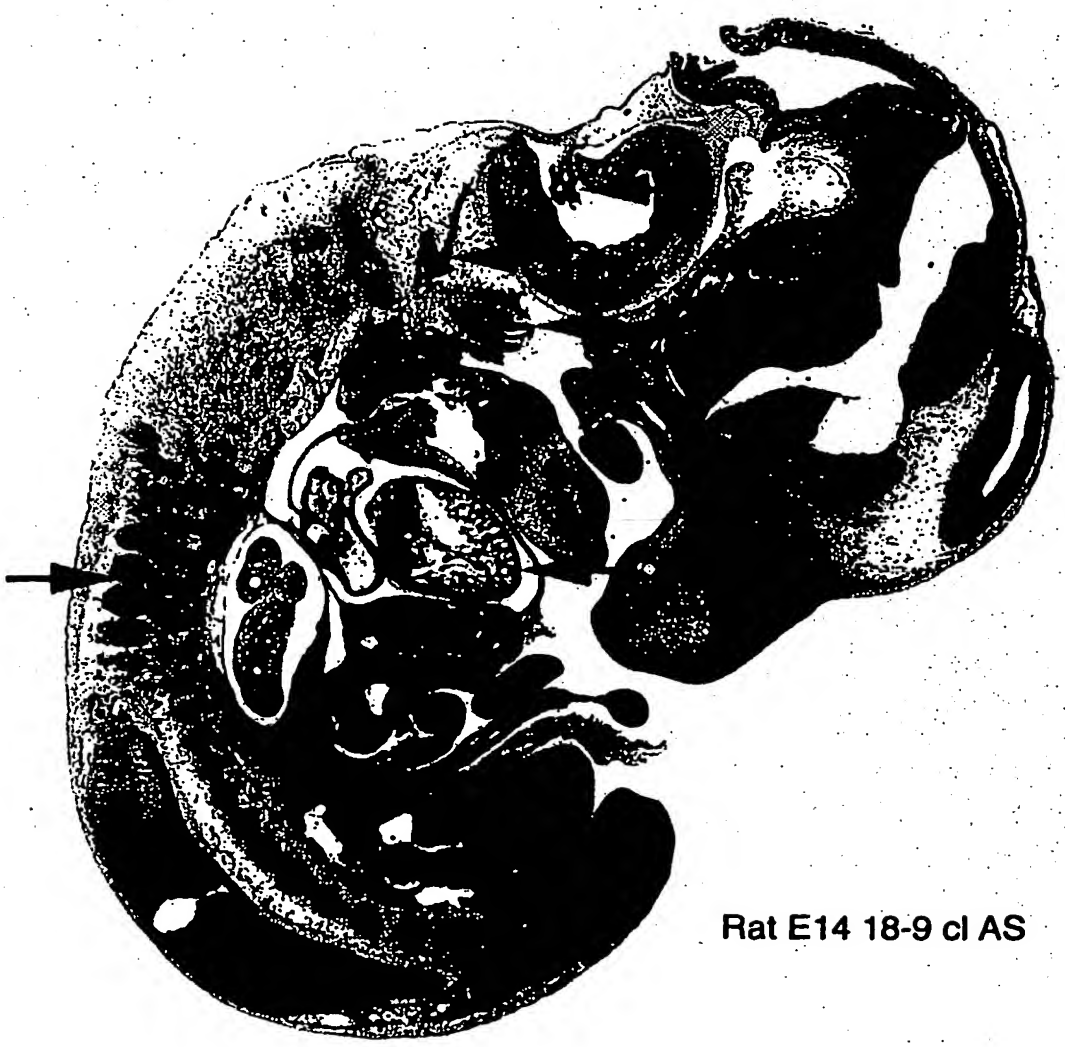

FIG. 3B

MOTIF STUDY IN KINASE DOMAIN THAT PREDICT KINASE SPECIFICITY

KINASE	SUBDOMAIN	FUNCTION
1.GXGXXG	I	ATP BINDING SITE
2.HRDLKSKN	VIB	IN MOST OF STK
3.HRDLAARN	VIB	IN MOST OF TK
4.GTKRYMAPE	VIII	IN MOST OF STK
5.XP(IV)(K/R)W(T/M)	VIII	IN MOST OF TK
6.DFG	VII	IN ALL OF PK

GXGXXG: ATP BINDING SITE NOT MATCH ON 9-2-1.9kb

TABLE 2 COMPARE WITH OTHER STKR MEMBERS

KINASE	SUBDOMAIN VIB	VIII
1.9-2-1.9kb	HRDLKPEN	GTPCWMAPE
2.ActR-II	HRDLKSKN	GTRRYMAPE
3.ActR-IIB	HRDFKSKN	GTRRYMAPE
4.TBR-II	HRDLKSSN	GTARYMAPE
5.ALK-I	HRDFKSRN	GTKRYMAPE

FIG. 4

SEQUENCE LISTING

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1989 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..1882

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
C CCC CTC GAG GTC GAC GGT ATC GAT AAG CTT GAT ATC GAA TTC CGA          46
  Pro Leu Glu Val Asp Gly Ile Asp Lys Leu Asp Ile Glu Phe Arg
    1               5                  10                  15

GGC TCG GCC TCG GGC TCC CTC CCC TCG TGG CAG GCT CCG GCG GCG GCG        94
Gly Ser Ala Ser Gly Ser Leu Pro Ser Trp Gln Ala Pro Ala Ala Ala
                20                  25                  30

GCG GCG GCG GCT GCG GCG GAC GGG GAG GGC GTG CGC CGG TCG AGA GGT       142
Ala Ala Ala Ala Ala Ala Asp Gly Glu Gly Val Arg Arg Ser Arg Gly
            35                  40                  45

GTG GGC GAC GAG GCT AGG GAA GTT TCA AGT GGA AGG TCG TCC GCC GGC       190
Val Gly Asp Glu Ala Arg Glu Val Ser Ser Gly Arg Ser Ser Ala Gly
        50                  55                  60

CGG CGC GTC CCC TCA CTC TCC TCC CGC AGC ATC ATG GCG GAG CCG AGC       238
Arg Arg Val Pro Ser Leu Ser Ser Arg Ser Ile Met Ala Glu Pro Ser
    65                  70                  75

GGC TCG CCC GTG CAC ATC CAG CTT CCC CAG CAG GCG GCC CCG GTG ACA       286
Gly Ser Pro Val His Ile Gln Leu Pro Gln Gln Ala Ala Pro Val Thr
 80                  85                  90                  95

GCC GCG GCG GCG GCT CCG GCG GCC GCG ACA TCA GCA CCG GCC CCG GCC       334
Ala Ala Ala Ala Ala Pro Ala Ala Ala Thr Ser Ala Pro Ala Pro Ala
                100                 105                 110

CCG GCC CCG GCG GCG CCC GCA GCC CCG GCC CCG GCT CCA GCT GCG GCT       382
Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Pro Ala Ala Ala
            115                 120                 125

CCA GCC CCG GCC CCG GCA GCT CAG GCG GTC GGC TGG CCC ATC TGC AGG       430
Pro Ala Pro Ala Pro Ala Ala Gln Ala Val Gly Trp Pro Ile Cys Arg
        130                 135                 140

GAC GCG TAC GAG CTC CAG GAG GTT ATC GGC AGT GGA GCG ACC GCC GTG       478
Asp Ala Tyr Glu Leu Gln Glu Val Ile Gly Ser Gly Ala Thr Ala Val
    145                 150                 155

GTT CAG GCA GCC CTG TGC AAA CCC AGG CAA GAA CGC GTA GCC ATA AAG       526
Val Gln Ala Ala Leu Cys Lys Pro Arg Gln Glu Arg Val Ala Ile Lys
160                 165                 170                 175

CGG ATC AAC TTG GAA AAG TGC CAG ACG AGT ATG GAT GAA CTC TTA AAA       574
```

```
Arg Ile Asn Leu Glu Lys Cys Gln Thr Ser Met Asp Glu Leu Leu Lys
                180                 185                 190

GAA ATT CAA GCC ATG AGC CAG TGC AGC CAT CCC AAC GTA GTG ACT TAT      622
Glu Ile Gln Ala Met Ser Gln Cys Ser His Pro Asn Val Val Thr Tyr
            195                 200                 205

TAT ACT TCC TTT GTG GTC AAA GAT GAA CTT TGG CTG GTC ATG AAA TTA      670
Tyr Thr Ser Phe Val Val Lys Asp Glu Leu Trp Leu Val Met Lys Leu
        210                 215                 220

CTA AGT GGA GGT TCC ATG TTG GAT ATC ATC AAA TAC ATC GTC AAT CGG      718
Leu Ser Gly Gly Ser Met Leu Asp Ile Ile Lys Tyr Ile Val Asn Arg
    225                 230                 235

GGA GAG CAT AAG AAT GGT GTC CTG GAA GAG GCA ATA ATC GCA ACA ATC      766
Gly Glu His Lys Asn Gly Val Leu Glu Glu Ala Ile Ile Ala Thr Ile
240                 245                 250                 255

CTT AAG GAG GTT TTG GAA GGA TTA GAC TAT CTG CAT AGA AAT GGT CAG      814
Leu Lys Glu Val Leu Glu Gly Leu Asp Tyr Leu His Arg Asn Gly Gln
                260                 265                 270

ATC CAT AGG GAT TTG AAA GCT GGC AAT ATT CTT CTG GGT GAG GAT GGA      862
Ile His Arg Asp Leu Lys Ala Gly Asn Ile Leu Leu Gly Glu Asp Gly
            275                 280                 285

TCA GTA CAG ATA GCA GAT TTT GGA GTA AGT GCA TTC TTA GCA ACA GGG      910
Ser Val Gln Ile Ala Asp Phe Gly Val Ser Ala Phe Leu Ala Thr Gly
        290                 295                 300

GGT GAT GTC ACA AGG AAT AAA GTC AGA AAA ACA TTT GTT GGT ACC CCA      958
Gly Asp Val Thr Arg Asn Lys Val Arg Lys Thr Phe Val Gly Thr Pro
    305                 310                 315

TGT TGG ATG GCC CCT GAG GTC ATA GAA CAG GTG AGA GGC TAT GAC TTC     1006
Cys Trp Met Ala Pro Glu Val Ile Glu Gln Val Arg Gly Tyr Asp Phe
320                 325                 330                 335

AAG GCT GAC ATG TGG AGT TTT GGA ATA ACA GCC ATT GAA TTA GCA ACG     1054
Lys Ala Asp Met Trp Ser Phe Gly Ile Thr Ala Ile Glu Leu Ala Thr
                340                 345                 350

GGA GCA GCG CCT TAC CAC AAA TAC CCT CCA ATG AAA GTG CTA ATG TTG     1102
Gly Ala Ala Pro Tyr His Lys Tyr Pro Pro Met Lys Val Leu Met Leu
            355                 360                 365

ACT TTG CAA AAT GAC CCG CCC ACT TTA GAA ACC GGC GTA GAG GAT AAA     1150
Thr Leu Gln Asn Asp Pro Pro Thr Leu Glu Thr Gly Val Glu Asp Lys
        370                 375                 380

GAA ATG ATG AAA AAA TAC GGC AAG TCC TTC AGA AAG TTA CTT TCA CTG     1198
Glu Met Met Lys Lys Tyr Gly Lys Ser Phe Arg Lys Leu Leu Ser Leu
    385                 390                 395

TGT CTC CAG AAA GAT CCT TCC AAA AGG CCC ACA GCA GCA GAA CTT TTA     1246
Cys Leu Gln Lys Asp Pro Ser Lys Arg Pro Thr Ala Ala Glu Leu Leu
400                 405                 410                 415

AAA TGC AAA TTC TTC CAG AAA GCC AAG AAC AGA GAG TAC CTG ATC GAG     1294
Lys Cys Lys Phe Phe Gln Lys Ala Lys Asn Arg Glu Tyr Leu Ile Glu
                420                 425                 430

AAG TTG CTG ACA CGA ACA CCA GAC ATA GCC CAA AGA GCC AAG AAG GTC     1342
Lys Leu Leu Thr Arg Thr Pro Asp Ile Ala Gln Arg Ala Lys Lys Val
            435                 440                 445
```

```
AGG CGA GTT CCT GGG TCA AGC GGT CAC CTT CAC AAG ACT GAA GAT GGC      1390
Arg Arg Val Pro Gly Ser Ser Gly His Leu His Lys Thr Glu Asp Gly
        450                 455                 460

GAC TGG GAG TGG AGT GAT GAT GAG ATG GAT GAG AAG AGT CAG GAG GGG      1438
Asp Trp Glu Trp Ser Asp Asp Glu Met Asp Glu Lys Ser Gln Glu Gly
    465                 470                 475

AAA GCG GCT GCC TCT CAA GAG AAG TCA CGA AGA GTA AAA GAA GAG AAC      1486
Lys Ala Ala Ala Ser Gln Glu Lys Ser Arg Arg Val Lys Glu Glu Asn
480                 485                 490                 495

CCA GAG ATC TCG GTG AAC GCT GGT GGC ATC CCC GAG CAA ATA CAG TCC      1534
Pro Glu Ile Ser Val Asn Ala Gly Gly Ile Pro Glu Gln Ile Gln Ser
                500                 505                 510

CTC TCC GTG CAC GAC TCT CAG GGC CAA CCA AAT GCT AAT GAA GAC TAC      1582
Leu Ser Val His Asp Ser Gln Gly Gln Pro Asn Ala Asn Glu Asp Tyr
            515                 520                 525

AGA GAA GGT CCT TGT GCG GTC AAC CTT GTT TTA AGA TTA AGA AAC TCC      1630
Arg Glu Gly Pro Cys Ala Val Asn Leu Val Leu Arg Leu Arg Asn Ser
        530                 535                 540

AGA AAG GAA CTT AAT GAC ATA CGA TTT GAG TTT ACT CCA GGA AGA GAT      1678
Arg Lys Glu Leu Asn Asp Ile Arg Phe Glu Phe Thr Pro Gly Arg Asp
    545                 550                 555

ACA GCA GAT GGT GTG TCT CAG GAG CTC TTC TCT GCT GGC TTG GTT GAC      1726
Thr Ala Asp Gly Val Ser Gln Glu Leu Phe Ser Ala Gly Leu Val Asp
560                 565                 570                 575

GGT CAT GAT GTA GTT ATA GTG GCT GCT AAT TTA CAG AAG ATT GTA GAT      1774
Gly His Asp Val Val Ile Val Ala Ala Asn Leu Gln Lys Ile Val Asp
                580                 585                 590

GAC CCC AAA GCT TTA AAA ACG TTG ACA TTT AAG TTG GCT TCT GGC TGT      1822
Asp Pro Lys Ala Leu Lys Thr Leu Thr Phe Lys Leu Ala Ser Gly Cys
            595                 600                 605

GAT GGG GCG GAG ATT CCT GAC GAA GTG AAG CTG ATC GGG TTC GCC CAG      1870
Asp Gly Ala Glu Ile Pro Asp Glu Val Lys Leu Ile Gly Phe Ala Gln
        610                 615                 620

TTG AGT GTG AGC TGATGTATGT CCCTTGCTGT CACCCTCATC TGTCGTGCGG          1922
Leu Ser Val Ser
    625

AATTCCTGCA GCCCGGGGGA TCCACTAGTT CTAGAGCGGC CGCCACCGCG GTGGAGCTCC    1982

AGCTTTT                                                             1989
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 627 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Pro Leu Glu Val Asp Gly Ile Asp Lys Leu Asp Ile Glu Phe Arg Gly
```

```
1               5                  10                  15
Ser Ala Ser Gly Ser Leu Pro Ser Trp Gln Ala Pro Ala Ala Ala Ala
            20                  25                  30
Ala Ala Ala Ala Ala Asp Gly Glu Gly Val Arg Arg Ser Arg Gly Val
        35                  40                  45
Gly Asp Glu Ala Arg Glu Val Ser Ser Gly Arg Ser Ser Ala Gly Arg
    50                  55                  60
Arg Val Pro Ser Leu Ser Ser Arg Ser Ile Met Ala Glu Pro Ser Gly
65                  70                  75                  80
Ser Pro Val His Ile Gln Leu Pro Gln Gln Ala Ala Pro Val Thr Ala
                85                  90                  95
Ala Ala Ala Ala Pro Ala Ala Ala Thr Ser Ala Pro Ala Pro Ala Pro
            100                 105                 110
Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Pro Ala Ala Ala Pro
        115                 120                 125
Ala Pro Ala Pro Ala Ala Gln Ala Val Gly Trp Pro Ile Cys Arg Asp
    130                 135                 140
Ala Tyr Glu Leu Gln Glu Val Ile Gly Ser Gly Ala Thr Ala Val Val
145                 150                 155                 160
Gln Ala Ala Leu Cys Lys Pro Arg Gln Glu Arg Val Ala Ile Lys Arg
                165                 170                 175
Ile Asn Leu Glu Lys Cys Gln Thr Ser Met Asp Glu Leu Leu Lys Glu
            180                 185                 190
Ile Gln Ala Met Ser Gln Cys Ser His Pro Asn Val Val Thr Tyr Tyr
        195                 200                 205
Thr Ser Phe Val Val Lys Asp Glu Leu Trp Leu Val Met Lys Leu Leu
    210                 215                 220
Ser Gly Gly Ser Met Leu Asp Ile Ile Lys Tyr Ile Val Asn Arg Gly
225                 230                 235                 240
Glu His Lys Asn Gly Val Leu Glu Glu Ala Ile Ile Ala Thr Ile Leu
                245                 250                 255
Lys Glu Val Leu Glu Gly Leu Asp Tyr Leu His Arg Asn Gly Gln Ile
            260                 265                 270
His Arg Asp Leu Lys Ala Gly Asn Ile Leu Leu Gly Glu Asp Gly Ser
        275                 280                 285
Val Gln Ile Ala Asp Phe Gly Val Ser Ala Phe Leu Ala Thr Gly Gly
    290                 295                 300
Asp Val Thr Arg Asn Lys Val Arg Lys Thr Phe Val Gly Thr Pro Cys
305                 310                 315                 320
Trp Met Ala Pro Glu Val Ile Glu Gln Val Arg Gly Tyr Asp Phe Lys
                325                 330                 335
Ala Asp Met Trp Ser Phe Gly Ile Thr Ala Ile Glu Leu Ala Thr Gly
            340                 345                 350
Ala Ala Pro Tyr His Lys Tyr Pro Pro Met Lys Val Leu Met Leu Thr
```

355 360 365

Leu Gln Asn Asp Pro Pro Thr Leu Glu Thr Gly Val Glu Asp Lys Glu
370 375 380

Met Met Lys Lys Tyr Gly Lys Ser Phe Arg Lys Leu Leu Ser Leu Cys
385 390 395 400

Leu Gln Lys Asp Pro Ser Lys Arg Pro Thr Ala Ala Glu Leu Leu Lys
405 410 415

Cys Lys Phe Phe Gln Lys Ala Lys Asn Arg Glu Tyr Leu Ile Glu Lys
420 425 430

Leu Leu Thr Arg Thr Pro Asp Ile Ala Gln Arg Ala Lys Lys Val Arg
435 440 445

Arg Val Pro Gly Ser Ser Gly His Leu His Lys Thr Glu Asp Gly Asp
450 455 460

Trp Glu Trp Ser Asp Asp Glu Met Asp Glu Lys Ser Gln Glu Gly Lys
465 470 475 480

Ala Ala Ala Ser Gln Glu Lys Ser Arg Arg Val Lys Glu Glu Asn Pro
485 490 495

Glu Ile Ser Val Asn Ala Gly Gly Ile Pro Glu Gln Ile Gln Ser Leu
500 505 510

Ser Val His Asp Ser Gln Gly Gln Pro Asn Ala Asn Glu Asp Tyr Arg
515 520 525

Glu Gly Pro Cys Ala Val Asn Leu Val Leu Arg Leu Arg Asn Ser Arg
530 535 540

Lys Glu Leu Asn Asp Ile Arg Phe Glu Phe Thr Pro Gly Arg Asp Thr
545 550 555 560

Ala Asp Gly Val Ser Gln Glu Leu Phe Ser Ala Gly Leu Val Asp Gly
565 570 575

His Asp Val Val Ile Val Ala Ala Asn Leu Gln Lys Ile Val Asp Asp
580 585 590

Pro Lys Ala Leu Lys Thr Leu Thr Phe Lys Leu Ala Ser Gly Cys Asp
595 600 605

Gly Ala Glu Ile Pro Asp Glu Val Lys Leu Ile Gly Phe Ala Gln Leu
610 615 620

Ser Val Ser
625

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 775 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TAAGCTTGCG TCCGCGACAG AGCGCCGGCC CCGGAAGCTC CCGCGGCCCC GGCCCCGGCC 60

CCGGCCCCGG CCCCGGCGGC ACAGGCTGTC GGCTGGCCCA TCTGCAGGGA CGCGTACGAG 120

CTGCAGGAGG TTATCGGCAG TGGAGCTACT GCTGTGGTTC AGGCAGCCCT ATGCAAACCC 180

AGGCAAGAAC GTGTAGCAAT AAAACGGATC AACTTGGAAA AATGCCAGAC CAGTATGGAT 240

GAACTACTTC TTGTGCCGTG AACCTCGTTT TGAGATTAAG AAACTCCAGA AAGGAACTTA 300

ATGACATACG ATTTGAGTTT ACTCCAGGAA GAGATACAGC AGATGGTGTA TCTCAGGAGC 360

TCTTCTCTGC TGGCTTGGTG GATGGTCACG ATGTAGTTAT AGTGGCTGCT AATTTACAGA 420

AGATTGTAGA TGATCCCAAA GCTTTAAAAA CATTGACATT TAAGTTGGCT TCTGGCTGTG 480

ATGGGTCGGA GATTCCTGAT GAAGTGAAGC TGATTGGGTT TGCTCAGTTG AGTGTCAGCT 540

GATGTATGTC CCTTGATGTC ACCCTGATCT GTCATGCCCC ACCGCCACCC CTACTCCTTC 600

AACCCTCCCT CTTTCTGCCC ATTTCCTCCC ACCCCCTCAC TCCCATTTCC TAGCAAAATC 660

AGAAGATTGT GAAGAGGCCG GCTTCAACAA AATGGGATAA AAAAATAATT TTTTAAAACT 720

TAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAA 775

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 520 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Ala Cys Val Arg Asp Arg Ala Pro Ala Pro Xaa Ala Pro Ala Ala Pro
1 5 10 15

Ala Pro Ala Pro Ala Pro Ala Pro Ala Ala Gln Ala Val Gly Trp Pro
20 25 30

Ile Cys Arg Asp Ala Tyr Glu Leu Gln Glu Val Ile Gly Ser Gly Ala
35 40 45

Thr Ala Val Val Gln Ala Ala Leu Cys Lys Pro Arg Gln Glu Arg Val
50 55 60

Ala Ile Lys Arg Ile Asn Leu Glu Lys Cys Gln Thr Ser Met Asp Glu
65 70 75 80

Leu Leu Lys Glu Ile Gln Ala Met Ser Gln Cys Ser His Pro Asn Val
85 90 95

Val Thr Tyr Tyr Thr Ser Phe Val Val Lys Asp Glu Leu Trp Leu Val
100 105 110

Met Lys Leu Leu Ser Gly Gly Ser Met Leu Asp Ile Ile Lys Tyr Ile
115 120 125

Val Asn Arg Gly Glu His Lys Asn Gly Val Leu Glu Glu Ala Ile Ile
130 135 140

Ala Thr Ile Leu Lys Glu Val Leu Glu Gly Leu Asp Tyr Leu His Arg
145 150 155 160

Asn Gly Gln Ile His Arg Asp Leu Lys Ala Gly Asn Ile Leu Leu Gly
165 170 175

Glu Asp Gly Ser Val Gln Ile Ala Asp Phe Gly Val Ser Ala Phe Leu
180 185 190

Ala Thr Gly Gly Asp Val Thr Arg Asn Lys Val Arg Lys Thr Phe Val
195 200 205

Gly Thr Pro Cys Trp Met Ala Pro Glu Val Met Glu Gln Val Arg Gly
210 215 220

Tyr Asp Phe Lys Ala Asp Met Trp Ser Phe Gly Ile Thr Ala Ile Glu
225 230 235 240

Leu Ala Thr Gly Ala Ala Pro Tyr His Lys Tyr Pro Pro Met Lys Val
245 250 255

Leu Met Leu Thr Leu Gln Asn Asp Pro Pro Thr Leu Glu Thr Gly Val
260 265 270

Glu Asp Lys Glu Met Met Lys Lys Tyr Gly Lys Ser Phe Arg Lys Leu
275 280 285

Leu Ser Leu Cys Leu Gln Lys Asp Pro Ser Lys Arg Pro Thr Ala Ala
290 295 300

Glu Leu Leu Lys Cys Lys Phe Phe Gln Lys Ala Lys Asn Arg Glu Tyr
305 310 315 320

Leu Ile Glu Lys Leu Leu Thr Arg Thr Pro Asp Ile Ala Gln Arg Ala
325 330 335

Lys Lys Val Arg Arg Val Pro Gly Ser Ser Gly His Leu His Lys Thr
340 345 350

Glu Asp Gly Asp Trp Glu Trp Ser Asp Asp Glu Met Asp Glu Lys Ser
355 360 365

Glu Glu Gly Lys Ala Ala Phe Ser Gln Glu Lys Ser Arg Arg Val Lys
370 375 380

Glu Glu Asn Pro Glu Ile Ala Val Ser Ala Ser Thr Ile Pro Glu Gln
385 390 395 400

Ile Gln Ser Leu Ser Val His Asp Ser Gln Gly Pro Pro Asn Ala Asn
405 410 415

Glu Asp Tyr Arg Glu Ala Ser Ser Cys Ala Val Asn Leu Val Leu Arg
420 425 430

Leu Arg Asn Ser Arg Lys Glu Leu Asn Asp Ile Arg Phe Glu Phe Thr
435 440 445

Pro Gly Arg Asp Thr Ala Asp Gly Val Ser Gln Glu Leu Phe Ser Ala
450 455 460

Gly Leu Val Asp Gly His Asp Val Val Ile Val Ala Ala Asn Leu Gln
465 470 475 480

Lys Ile Val Asp Asp Pro Lys Ala Leu Lys Thr Leu Thr Phe Lys Leu
485 490 495

Ala Ser Gly Cys Asp Gly Ser Glu Ile Pro Asp Glu Val Lys Leu Ile
500 505 510

Gly Phe Ala Gln Leu Ser Val Ser
515 520

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3056 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GCTTGCGTCC GCGACAGAGC GCCGGCCCCG GAAGCTCCCG CGGCCCCGGC CCCGGCCCCG 60

GCCCCGGCCC CGGCGGCACA GGCTGTCGGC TGGCCCATCT GCAGGGACGC GTACGAGCTG 120

CAGGAGGTTA TCGGCAGTGG AGCTACTGCT GTGGTTCAGG CAGCCCTATG CAAACCCAGG 180

CAAGAACGTG TAGCAATAAA ACGGATCAAC TTGGAAAAAT GCCAGACCAG TATGGATGAA 240

CTATTAAAAG AAATTCAAGC CATGAGTCAG TGCAGCCATC CCAACGTAGT GACCTATTAC 300

ACCTCTTTTG TGGTCAAAGA TGAACTTTGG CTGGTCATGA AATTACTAAG TGGAGGTTCA 360

ATGTTGGATA TCATAAAATA CATTGTCAAC CGAGGAGAAC ACAAGAATGG AGTTCTGGAA 420

GAGGCAATAA TAGCAACAAT TCTTAAAGAG GTTTTGGAAG GCTTAGACTA TCTACACAGA 480

AACGGTCAGA TTCACAGGGA TTTGAAAGCT GGTAATATTC TTCTGGGTGA GGATGGTTCA 540

GTACAAATAG CAGATTTTGG GGTAAGTGCG TTCCTAGCAA CAGGGGGTGA TGTTACCCGA 600

AATAAAGTAA GAAAAACATT CGTTGGCACC CCATGTTGGA TGGCTCCTGA AGTCATGGAA 660

CAGGTGAGAG GCTATGACTT CAAGGCTGAC ATGTGGAGTT TTGGAATAAC TGCCATTGAA 720

TTAGCAACAG GAGCAGCGCC TTATCACAAA TATCCTCCCA TGAAAGTGTT AATGTTGACT 780

TTGCAAAATG ATCCACCCAC TTTGGAAACA GGGGTAGAGG ATAAAGAAAT GATGAAAAAG 840

TACGGCAAGT CCTTTAGAAA ATTACTTTCA CTGTGTCTTC AGAAAGATCC TTCCAAAAGG 900

CCCACAGCAG CAGAACTTTT AAAATGCAAA TTCTTCCAGA AAGCCAAGAA CAGAGAGTAC 960

CTGATTGAGA AGCTGCTTAC AAGAACACCA GACATAGCCC AAAGAGCCAA AAAGGTAAGA 1020

AGAGTTCCTG GGTCAAGTGG TCACCTTCAT AAAACCGAAG ACGGGGACTG GGAGTGGAGT 1080

GACGACGAGA TGGATGAGAA GAGCGAAGAA GGGAAAGCAG CTTTTTCTCA GGAAAAGTCA 1140

CGAAGAGTAA AAGAAGAAAA TCCAGAGATT GCAGTGAGTG CCAGCACCAT CCCCGAACAA 1200

ATACAGTCCC TCTCTGTGCA CGACTCTCAG GGCCCACCCA ATGCTAATGA AGACTACAGA 1260

GAAGCTTCTT CTTGTGCCGT GAACCTCGTT TTGAGATTAA GAAACTCCAG AAAGGAACTT 1320

AATGACATAC GATTTGAGTT TACTCCAGGA AGAGATACAG CAGATGGTGT ATCTCAGGAG 1380

CTCTTCTCTG CTGGCTTGGT GGATGGTCAC GATGTAGTTA TAGTGGCTGC TAATTTACAG 1440

AAGATTGTAG ATGATCCCAA AGCTTTAAAA ACATTGACAT TTAAGTTGGC TTCTGGCTGT 1500

GATGGGTCGG AGATTCCTGA TGAAGTGAAG CTGATTGGGT TTGCTCAGTT GAGTGTCAGC 1560

TGATGTATGT CCCTTGATGT CACCCTGATC TGTCATGCCC CACCGCCACC CCTACTCCTT 1620

CAACCCTCCC TCTTTCTGCC CATTTCCTCC CACCCCCTCA CTCCCATTTC CTAGCAAAAT 1680

CAGAAGATTG TGAAGAGGCC GGCTTCAACA AAATGGGATA AAAAAATAAT TTTTTAAAAC 1740

TTACAACACT CCGAGTTCTG CTTTATTCTC TAGCAATCCA CAGTACAAGA ACAAGCAAAT 1800

GCCACAGCTG CACGACTGTT GCTCATTTTT CCAAAAGCTA TTTAATATTC TTAGCAATCA 1860

ATTTGGATAT CCCTTAAGTG AAAAGAATCT GAAATACACT CAGGTGGTCT TATTTATTGG 1920

CAACAAAAGG AATTTTCTAT CCAGAAGCCT ATTTCTCCTT TCATTGTTGT TATTTCTGTT 1980

ATAATACTTT AATTGTACAT CTGACAATAC TGCCTCTTTT ATGTTGTATT TAGAAATTAA 2040

TATACTTATA AAATTAAGAT TTATTAGCCA AACTTGAATT CTAGTTTTAA AACTGACTGT 2100

GAATTTTATT TTTCATATAT TTATGCATTA CACACCTTAG CTATAAGAAA AAAAGGGTTT 2160

TGATTATATG CTTCTTGCAG TTAATCTCGT TATTTAAACA AAAAGTTTTG GGTCTATCTT 2220

TGGAGTATTT GTAACTTCTA AATTTTGAAA TGACTGAATT AGGAATTTGG ATGCTTATTC 2280

TTTTAGTCTG TTTGCCTAAA AACCAATTTA CAATCTGACT GTCTCTTGGG AGAGGGAGGT 2340

GCCTTGCAAA CTTTCACATT AAGAATGTGC CTGAGGCTGC TTTACTCTGG AATAGTCTCA 2400

GATCTAAAAT TTCCTCTATA TAAGGTGGCA TATGTTAAGT TTTGCTTCAT TGGACCGTTT 2460

AGAATGCTAT GTAAAATGTT GCCATTCTGT TAGATTGCTA ACTATATACC CATCTCTGAT 2520

TTGGCTCTCC TTAAGTGATA GGATTTGTTA TTCTAAAGGT GATAAACTTG AAAATATCAG 2580

AATCTGAGTT TTACTTGAAA TTTTGCAGAA TACCCAGGTG GAGTGAAAAT TGGAAGGGTT 2640

TTGTGCAATG ACTAAAAGGT AAAACGCTGT TAAGGTTCAA GAATCAATAC TTTCAACCCA 2700

AGTAGCCCTC TGCTTGACTG TATATTATGG AACTAGTAAA CCTTAGGATT TTGAAAATTG 2760

GAGTCTAATC TTTCAAGGAG GTGGGCTCCC AGGATGGTAC CATTGCTCTT TCCTAGCTAA 2820

CCCTAGATAT GGCAGCTCTT TAATGTACTT CAAAAAGCAA ATATATATTA CTAAGGAAAA 2880

AAAGTTATTT ATAATTGCCT TGTCATAATT GTTAAGGTGT TCTAGAGCCA TTTGCATACA 2940

ATTTAATGTA ATTTCATTCC ATTCTATTGT TTACACAACG ATTACTCGAA GATGACTGCA 3000

AAGGTAAAAG GAAAATAAAA GTGTATTGCA CAATGAAAAA AAAAAAAAAA AAAAAA 3056

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 520 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Ala Cys Val Arg Asp Arg Ala Pro Ala Pro Glu Ala Pro Ala Ala Pro
 1               5                  10                  15

Ala Pro Ala Pro Ala Pro Ala Pro Ala Ala Gln Ala Val Gly Trp Pro
            20                  25                  30

Ile Cys Arg Asp Ala Tyr Glu Leu Gln Glu Val Ile Gly Ser Gly Ala
        35                  40                  45

Thr Ala Val Val Gln Ala Ala Leu Cys Lys Pro Arg Gln Glu Arg Val
    50                  55                  60

Ala Ile Lys Arg Ile Asn Leu Glu Lys Cys Gln Thr Ser Met Asp Glu
65                  70                  75                  80

Leu Leu Lys Glu Ile Gln Ala Met Ser Gln Cys Ser His Pro Asn Val
                85                  90                  95

Val Thr Tyr Tyr Thr Ser Phe Val Val Lys Asp Glu Leu Trp Leu Val
            100                 105                 110

Met Lys Leu Leu Ser Gly Gly Ser Met Leu Asp Ile Ile Lys Tyr Ile
        115                 120                 125

Val Asn Arg Gly Glu His Lys Asn Gly Val Leu Glu Glu Ala Ile Ile
    130                 135                 140

Ala Thr Ile Leu Lys Glu Val Leu Glu Gly Leu Asp Tyr Leu His Arg
145                 150                 155                 160

Asn Gly Gln Ile His Arg Asp Leu Lys Ala Gly Asn Ile Leu Leu Gly
                165                 170                 175

Glu Asp Gly Ser Val Gln Ile Ala Asp Phe Gly Val Ser Ala Phe Leu
            180                 185                 190

Ala Thr Gly Gly Asp Val Thr Arg Asn Lys Val Arg Lys Thr Phe Val
        195                 200                 205

Gly Thr Pro Cys Trp Met Ala Pro Glu Val Met Glu Gln Val Arg Gly
    210                 215                 220

Tyr Asp Phe Lys Ala Asp Met Trp Ser Phe Gly Ile Thr Ala Ile Glu
225                 230                 235                 240

Leu Ala Thr Gly Ala Ala Pro Tyr His Lys Tyr Pro Pro Met Lys Val
                245                 250                 255

Leu Met Leu Thr Leu Gln Asn Asp Pro Pro Thr Leu Glu Thr Gly Val
            260                 265                 270

Glu Asp Lys Glu Met Met Lys Lys Tyr Gly Lys Ser Phe Arg Lys Leu
        275                 280                 285

Leu Ser Leu Cys Leu Gln Lys Asp Pro Ser Lys Arg Pro Thr Ala Ala
    290                 295                 300
```

```
Glu Leu Leu Lys Cys Lys Phe Phe Gln Lys Ala Lys Asn Arg Glu Tyr
305                 310                 315                 320

Leu Ile Glu Lys Leu Leu Thr Arg Thr Pro Asp Ile Ala Gln Arg Ala
                325                 330                 335

Lys Lys Val Arg Arg Val Pro Gly Ser Ser Gly His Leu His Lys Thr
            340                 345                 350

Glu Asp Gly Asp Trp Glu Trp Ser Asp Asp Glu Met Asp Glu Lys Ser
        355                 360                 365

Glu Glu Gly Lys Ala Ala Phe Ser Gln Glu Lys Ser Arg Arg Val Lys
    370                 375                 380

Glu Glu Asn Pro Glu Ile Ala Val Ser Ala Ser Thr Ile Pro Glu Gln
385                 390                 395                 400

Ile Gln Ser Leu Ser Val His Asp Ser Gln Gly Pro Pro Asn Ala Asn
                405                 410                 415

Glu Asp Tyr Arg Glu Ala Ser Ser Cys Ala Val Asn Leu Val Leu Arg
            420                 425                 430

Leu Arg Asn Ser Arg Lys Glu Leu Asn Asp Ile Arg Phe Glu Phe Thr
        435                 440                 445

Pro Gly Arg Asp Thr Ala Asp Gly Val Ser Gln Glu Leu Phe Ser Ala
    450                 455                 460

Gly Leu Val Asp Gly His Asp Val Val Ile Val Ala Ala Asn Leu Gln
465                 470                 475                 480

Lys Ile Val Asp Asp Pro Lys Ala Leu Lys Thr Leu Thr Phe Lys Leu
                485                 490                 495

Ala Ser Gly Cys Asp Gly Ser Glu Ile Pro Asp Glu Val Lys Leu Ile
            500                 505                 510

Gly Phe Ala Gln Leu Ser Val Ser
        515                 520
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1776 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
GCCCTGCCCT GGTCCATCAA CAGGGACGAT TACGAGCTGC AGGAGGTGAT CGGGAGTGGA      60
GCAACTGCTG TAGTCCAAGC AGCTTATTGT GCCCCTAAAA AGGAGAAAGT GGCAATCAAA     120
CGGATAAACC TTGAGAAATG TCAAACTAGC ATGGATGAAC TCCTGAAAGA AATTCAAGCC     180
ATGAGTCAAT GCCATCATCC TAATATTGTA TCTTACTACA CATCTTTTGT GGTAAAAGAT     240
```

GAGCTGTGGC TTGTCATGAA GCTGCTAAGT GGAGGTTCTG TTCTGGATAT TATTAAGCAC 300

ATTGTGGCAA AAGGGGAACA CAAAAGTGGA GTCCTAGATG ANTCTACCAT TGCTACGATA 360

CTCCGAGAAG TACTGGAAGG GCTGGAATAT CTGCATAAAA NTGGACAGAT CCACAGAGAT 420

GTGAAAGCTG GAAACATTCT TNTTGGAGAA GATGGCTCAG TACAGATTGC AGACTTTGGG 480

GTTAGTGCTT TTTTAGCAAC TGGTGGTGAT ATTACCCGAA ATAAAGTGAG AAAGACCTTT 540

GTTGGCACCC CTTGTTGGAT GGCACCTGAA GTTATGGAAC AGGTCCGTGG TTATGATTTC 600

AAAGCTGATA TTTGGAGTTT TGGAATTACA GCAATTGAAT TGGCTACAGG GGCGGCTCCT 660

TATCATAAAT ATCCACCAAT GAAGGTTTTA ATGCTGACAC TGCAGAACGA TCCTCCTTCT 720

TTGGAAACTG GTGTTCAAGA TAAAGAAATG CTGAAAAAAT ATGGAAAATC ATTTAGAAAA 780

ATGATTTCAT TGTGCCTTCA AAAAGATCCA GAAAAAAGAC CAACAGCAGC AGAACTATTA 840

AGGCACAAAT TTTTCCAGAA AGCAAAGAAT AAAGAATTTC TTCAAGAAAA AACATTGCAG 900

AGAGCACCAA CCATTTCTGA AAGAGCAAAA AAGGTTCGGA GAGTACCAGG TTCCAGTGGG 960

CGTCTTCATA AGACAGAGGA TGGAGGCTGG GAGTGGAGTG ATGATGAATT TGATGAAGAA 1020

AGTGAGGAAG GGAAAGCAGC AATTTCACAA CTCAGGTCTC CCCGAGTGAA AGAATCAATA 1080

TCAAATTCTG AGCTCTTTCC AACAACTGAT CCTGTGGGTA CTTTGCTCCA AGTTCCAGAA 1140

CAGATCTCTG CTCATCTACC TCAGCCAGCT GGGCAGATTG CTACACAGCC AACTCAAGTC 1200

TCTCTCCCAC CCACCGCAGA GCCAGCAAAA ACAGCTCAGG CTTTGTCTTC AGGATCAGGT 1260

TCACAAGAAA CCAAGATCCC AATCAGTCTA GTACTAAGAT TAAGGAATTC CAAAAAAGAA 1320

CTAAATGATA TTCGATTTGA ATTTACTCCT GGGAGAGATA CAGCAGAGGG TGTCTCTCAG 1380

GAACTCATTT CTGCTGGCCT GGTCGACGGA AGGGATTTAG TAATAGTGGC AGCTAATTTG 1440

CAGAAAATTG TGGANGAACC TCAGTCAAAT CGATCTGTCA CTTTCAAACT GGCATCTGGT 1500

GTCGAAGGCT CAGATATTCC TGATGATGGT AAACTGATAG GATTTGCCCA GCTCAGCATC 1560

AGCTAAACCA CAACCCTGGA AGAGGCGGCC TAAGGAGATT CCACACATGC GTATCTCTGT 1620

NGCTTCTATT GGCCTAAACC CACTACTGCC AAAGAACCCA GCAACAAACC TCCCGGCTAG 1680

GAGCTTTAGA AGTCTTTATG TCTTCCTGCC ATCATTCCTC CTTTTCCCAC AGGGAAGAAA 1740

AGTTGGTCAC TAGTGGCCAG ATCCCAGAGT CCGTAG 1776

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1718 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TGGCACGAGG TTTTCAAATA TCAAATTGAA GAAAAACTCA ATCCTCCCAT GAAAATCAGT 60

TCGCCTGGCC TCCAAGTNGT GAGGAAATGG GTATGCAAGG CTGAGATTTC TACAGCAATA 120

AAGGAGACAC ACACTGGGCC AGAGAGGCCT GCCTTCTGCC TGCTCTCCTG CACTGACCCT 180

TTGGAGGGGG TCTCTGTGTG CTGAAGCTAA CTCAAGATGG AAAGTGAAAC CACATGTGCC 240

GTGACCTTTA GGTTTTATGA GTAGACAGTG TTCATTTGAT TTTCTACAGA AATAATATAA 300

ATTATTCTTT AGGTTTAAAA AAGAGCACTC ATAATGCAAT ATGTGAATAA TCAGTGAGGT 360

TGATTTTTCT TTTTTCCTAC CGTTTCATAG TCTTTGTCTA ACTGCTAGTA ACCCTACCGA 420

GTTTTATATA TGAGTGGGAT ACTCAATCTG GCCTTAAAAA GATACACAAA GATGGGCTGT 480

GGGTCCCTGG AAAGGGGGAG AGTTGCCCTT TACAGAATCA CTCGAGCCCT TTCCAGCACT 540

GTTGGTCTGA TGAACAAGGT TGTTTTACCT TATTTTCTCT TGGAACATAT CTGAAAACCT 600

TCCCCACAAA TAACTTGTCA CACCTTTTGT TTCATTCTGA GTCTTTAGTT TTAGTCATGG 660

GCTTTCTTCA CCTGCTCTAG GTGCAAAGGC ATGTTGGGAA AGAGATGGAT GTTGGGGAGG 720

AAGAGAGGAG ATGGATTTCA GTTGGGAGTT AGGACGAGAG TAGGTGAGAT GATCAGACAC 780

CGGAGTTCAA CGTCCCAGCA GTCTTGGTAA AAGGAGGGAG CCTGCTGAGC CAGGAGGGAG 840

AAAAGAAGAT TGACCAGCTT GCTAGAAAAA TACTTAGCTT TTCTTTTTCT TTTTTTGTSG 900

AGGGGGGACG GAGAGGAACA AGGATGGGGA GGTAGGAATG AGGTATAGAA AAGAGATAGC 960

ATCTTCTTTG GCACAAGACT AGTGGCTTAC CGCTTACCTT AGAGTTTTGT TTTTTTTTTT 1020

TCAAACCCAT CAAAATCTAC TTATTTATGA ATCCAAGGGG TGGCAGCATC ACTCTGTTCT 1080

AGCATTCTTT GTGGAGATGG TCTGGTGCCT AGCTGGGAGT GAGCAGCAGC CCATCCCCTG 1140

TTCACTTTCT CTAGCCCATC ATTACCTGTG AACTGCAGTG GGGCAGTCAC TTCTTGTGCC 1200

GTGAACCTCG TTTTGAGATT AAGAAACTCC AGAAAGGAAC TTAATGACAT ACGATTTGAG 1260

TTTACTCCAG GAAGAGATAC AGCAGATGGT GTATCTCAGG AGCTCTTCTC TGCTGGCTTG 1320

GTGGATGGTC ACGATGTAGT TATAGTGGCT GCTAATTTAC AGAAGATTGT AGATGATCCC 1380

AAAGCTTTAA AAACATTGAC ATTTAAGTTG GCTTCTGGCT GTGATGGGTC GGAGATTCCT 1440

GATGAAGTGA AGCTGATTGG GTTTGCTCAG TTGAGTGTCA GCTGATGTAT GTCCCTTGAT 1500

GTCACCCTGA TCTGTCATGC CCCACCGCCA CCCCTACTCC TTCAACCCTC CCTCTTTCTG 1560

CCCATTTCCT CCCACCCCCT CACTCCCATT TCCTAGCAAA ATCAGAAGAT TGTGAAGAGG 1620

CCGGCTTCAA CAAAATGGGA TAAAAAAATA ATTTTTTAAA ACTTAAAAAA AAAAAAAAAA 1680

AAAAAAAAAA AAAAAAAAAA AAAAAAAA 1718

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 469 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
Ala Leu Pro Trp Ser Ile Asn Arg Asp Asp Tyr Glu Leu Gln Glu Val
1               5                  10                  15

Ile Gly Ser Gly Ala Thr Ala Val Val Gln Ala Ala Tyr Cys Ala Pro
            20                  25                  30

Lys Lys Glu Lys Val Ala Ile Lys Arg Ile Asn Leu Glu Lys Cys Gln
        35                  40                  45

Thr Ser Met Asp Glu Leu Leu Lys Glu Ile Gln Ala Met Ser Gln Cys
    50                  55                  60

His His Pro Asn Ile Val Ser Tyr Tyr Thr Ser Phe Val Val Lys Asp
65                  70                  75                  80

Glu Leu Trp Leu Val Met Lys Leu Leu Ser Gly Gly Ser Val Leu Asp
                85                  90                  95

Ile Ile Lys His Ile Val Ala Lys Gly Glu His Lys Ser Gly Val Leu
            100                 105                 110

Asp Xaa Ser Thr Ile Ala Thr Ile Leu Arg Glu Val Leu Glu Gly Leu
        115                 120                 125

Glu Tyr Leu His Lys Xaa Gly Gln Ile His Arg Asp Val Lys Ala Gly
    130                 135                 140

Asn Ile Leu Xaa Gly Glu Asp Gly Ser Val Gln Ile Ala Asp Phe Gly
145                 150                 155                 160

Val Ser Ala Phe Leu Ala Thr Gly Gly Asp Ile Thr Arg Asn Lys Val
                165                 170                 175

Arg Lys Thr Phe Val Gly Thr Pro Cys Trp Met Ala Pro Glu Val Met
            180                 185                 190

Glu Gln Val Arg Gly Tyr Asp Phe Lys Ala Asp Ile Trp Ser Phe Gly
        195                 200                 205

Ile Thr Ala Ile Glu Leu Ala Thr Gly Ala Ala Pro Tyr His Lys Tyr
    210                 215                 220

Pro Pro Met Lys Val Leu Met Leu Thr Leu Gln Asn Asp Pro Pro Ser
225                 230                 235                 240

Leu Glu Thr Gly Val Gln Asp Lys Glu Met Leu Lys Lys Tyr Gly Lys
                245                 250                 255

Ser Phe Arg Lys Met Ile Ser Leu Cys Leu Gln Lys Asp Pro Glu Lys
            260                 265                 270

Arg Pro Thr Ala Ala Glu Leu Leu Arg His Lys Phe Phe Gln Lys Ala
        275                 280                 285

Lys Asn Lys Glu Phe Leu Gln Glu Lys Thr Leu Gln Arg Ala Pro Thr
```

290 295 300

Ile Ser Glu Arg Ala Lys Lys Val Arg Arg Val Pro Gly Ser Ser Gly
305 310 315 320

Arg Leu His Lys Thr Glu Asp Gly Gly Trp Glu Trp Ser Asp Asp Glu
325 330 335

Phe Asp Glu Glu Ser Glu Glu Gly Lys Ala Ala Ile Ser Gln Leu Arg
340 345 350

Ser Pro Arg Val Lys Glu Ser Ile Ser Asn Ser Glu Leu Phe Pro Thr
355 360 365

Thr Asp Pro Val Gly Thr Leu Leu Gln Val Pro Glu Gln Ile Ser Ala
370 375 380

His Leu Pro Gln Pro Ala Gly Gln Ile Ala Thr Gln Pro Thr Gln Val
385 390 395 400

Ser Leu Pro Pro Thr Ala Glu Pro Ala Lys Thr Ala Gln Ala Leu Ser
405 410 415

Ser Gly Ser Gly Ser Gln Glu Thr Lys Ile Pro Ile Ser Leu Val Leu
420 425 430

Arg Leu Arg Asn Ser Lys Lys Glu Leu Asn Asp Ile Arg Phe Glu Phe
435 440 445

Thr Pro Gly Arg Asp Thr Ala Glu Gly Val Ser Gln Glu Leu Ile Ser
450 455 460

Ala Gly Leu Val Asp
465

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/54, 9/12, C07K 16/40, C12Q 1/68, A61K 38/45, A01K 67/027, G01N 33/50, 33/573

A3

(11) International Publication Number: WO 99/07854

(43) International Publication Date: 18 February 1999 (18.02.99)

(21) International Application Number: PCT/US98/16640

(22) International Filing Date: 11 August 1998 (11.08.98)

(30) Priority Data:
08/909,354 11 August 1997 (11.08.97) US

(71) Applicant (*for all designated States except US*): ONTOGENY, INC. [US/US]; 45 Moulton Street, Cambridge, MA 02138 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): PANG, Kevin [US/US]; 53 Hill Road #10, Belmont, MA 02178 (US). MIAO, Ningning [US/US]; 671 Washington Street, Brookline, MA 02146 (US). BARKER, Douglas, D. [US/US]; 47 Louise Road, Belmont, MA (US).

(74) Agents: VINCENT, Matthew, P. et al.; Foley, Hoag & Eliot, LLP, One Post Office Square, Boston, MA 02109 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
15 April 1999 (15.04.99)

(54) Title: SERINE/THREONINE KINASE, AND USES RELATED THERETO

MOTIF STUDY IN KINASE DOMAIN THAT PREDICT KINASE SPECIFICITY

KINASE	SUBDOMAIN	FUNCTION
1.GXGXXG	I	ATP BINDING SITE
2.HRDLKSKN	VIB	IN MOST OF STK
3.HRDLAARN	VIB	IN MOST OF TK
4.GTKRYMAPE	VIII	IN MOST OF STK
5.XP(IV)(K/R)W(T/M)	VIII	IN MOST OF TK
6.DFG	VII	IN ALL OF PK

GXGXXG: ATP BINDING SITE NOT MATCH ON 9-2-1.9kb

TABLE 2 COMPARE WITH OTHER STKR MEMBERS

KINASE	SUBDOMAIN VIB	VIII
1.9-2-1.9kb	HRDLKPEN	GTPCWMAPE
2.ActR-II	HRDLKSKN	GTRRYMAPE
3.ActR-IIB	HRDFKSKN	GTRRYMAPE
4.TBR-II	HRDLKSSN	GTARYMAPE
5.ALK-I	HRDFKSRN	GTKRYMAPE

(57) Abstract

We describe here a new class of serine/threonine kinase receptors, called "*pan-s/tk*". The sequence of exemplary *pan-s/tk* genes indicates that they encode receptor type serine/threonine kinases with a single kinase domain.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

INTERNATIONAL SEARCH REPORT

Inte 'onal Application No
PCT/US 98/16640

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/54 C12N9/12 C07K16/40 C12Q1/68 A61K38/45
A01K67/027 G01N33/50 G01N33/573

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K C12Q A61K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category '	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	M.D. ADAMS ET AL.: "Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequences" EMBL SEQUENCE DATABASE,18 April 1997, XP002091521 Heidelberg, FRG EST82343 Prostate gland I Homo sapiens cDNA 5' end similar to serine/threonine kinase; Accession no. AA370646; --- -/--	19-21,30

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 January 1999	18/02/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hornig, H

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/16640

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category '	Citation of document, with indication,where appropriate. of the relevant passages	Relevant to claim No.
X	"National Cancer Institute, Cancer Genome Anatomy Project (CGAP), " EMBL SEQUENCE DTABASE,12 April 1997, XP002091522 Heidelberg, FRG zs55g10.r1 NCI_CGAP_CGB1 Homo sapiens cDNA clone IMAGE:701442 5' similar to TR:E225586 E225586 Serine-Threonine proteine kinase, Accession no. AA286878; ---	30
P,X	H. USHIRO ET AL.: "Molecular cloning and characterization of a novel Ste20-related protein kinase enriched in neurons and transporting epithelia" ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS, vol. 355, no. 2, 15 July 1998, pages 233-240, XP002091523 ACADEMIC PRESS, INC., NY, US see the whole document ---	1-31
P,X	N. MIAO ET AL.: "PS/TK, a novel serine/threonine kinase expressed during pancreatic development" EMBL SEQUENCE DATABASE,17 June 1998, XP002091524 Heidelberg, FRG Accession no. AF068261 ---	1-31
P,X	D. BAYTEL AND J. DON : "Homo sapiens DCHT mRNA, complete cds." EMBL SEQUENCE DATABASE,23 September 1997, XP002091525 Heidelberg, FRG Accession no. AF017635 ---	1-31
A	CREASY C L ET AL: "Cloning and characterization of a member of the MST subfamily of Ste20-like kinases" GENE, vol. 167, no. 1, 1995, page 303-306 XP004043060 see the whole document ---	1-44
A	CRASY C L ET AL: "CLONING AND CHARACTERIZATION OF A HUMAN PROTEIN KINASE WITH HOMOLOGY TO STE20" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 270, no. 37, 15 September 1995. pages 21695-21700, XP002054149 see the whole document ---	1-44

-/--

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 98/16640

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SUSUMU ITOH ET AL: "Molecular cloning and characterization of a novel putative STE20 -like kinase in Guinea pigs" ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS, vol. 340, no. 2, 15 April 1997, pages 201-207, XP002079735 see the whole document ---	1-44
A	POMBO ET AL: "Activation of a human Ste20-like kinase by oxidant stress defines a novel stress response pathway" EMBO JOURNAL, vol. 15, no. 17, 1996, pages 4537-4546, XP002086763 see the whole document -----	1-44

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

I. .national application No.

PCT/US 98/ 16640

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Remark: Although claims 34,35 and 38,41 (as far as in vivo methods are concerned) are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.